Folding and Stability of Homooligomeric Coiled Coils

Dissertation
zur Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät der
Universität Zürich

von
Saša Bjelić
aus
Wilchingen SH

Promotionskomitee

Prof. Dr. Amedeo Caflisch (Vorsitz)
PD Dr. Ilian Jelezarov (Leitung der Dissertation)
N. N.

Zürich 2008

To my lovely wife
and children

CONTENTS

Summary

Notwithstanding the rapid developments and important achievements in the areas of structural biology and protein biophysics in the past decades, many aspects of the protein folding problem remain unresolved. The basic reason is the conceptual complexity of the problem how does the unfolded polypeptide chain acquire a unique 3D conformation. Proteins fold on a broad and rough, multidimensional energetic landscape as the consequence of energetic factors, which are large in magnitude and are effectively compensating each other. Folding involves formation of a multitude of non-covalent bonds, dramatic restriction of degrees of freedom, and massive dehydration of molecular surface. Given the size of proteins, their structural inhomogeneity, and their ensemble nature, both theoretical methods and experiments still cannot capture details of the energetic landscape and the mechanistic sequence of events on the folding pathway.

Most of the current knowledge on protein folding stems from studies of small globular proteins or globular domains of larger multi-domain proteins. On the other hand, oligomeric proteins consisting of two or more non-covalently associated polypeptide chains are in fact more abundant in nature. Oligomers have appeared as an evolutionary solution reconciling the pressure of physical and chemical factors with diversification of biological function. Monomeric and oligomeric proteins are stabilized by the same forces and thermodynamic mechanisms. However, folding of oligomeric proteins is peculiar in two respects. First, in the succession of molecular events taking place between the unfolded and folded state in the course of folding, there is at least one obligatory step involving change in molecularity. Second, the ubiquitous energetic mechanisms stabilizing proteins operate simultaneously on the level of subunits and on the level of the oligomer, meaning that there must be a balance between inter-subunit and *intra*-subunit molecular interactions. As the consequence, oligomeric proteins exhibit highly variable folding pathways. Much more work is required to explore the details in folding of oligomeric proteins, in order to deepen our understanding of protein folding in general, and to utilize their great potential in biomedical and biotechnological applications.

The present thesis is devoted to biophysical characterization of three oligomeric proteins. They all share a common structural motif an α-helical coiled coil. For our investigations we on purpose selected proteins which contain coiled coils but exhibit structural or biophysical peculiarities.

The protein moiety of the *E. coli* outer membrane protein (Lpp-56) is a seven-heptades long trimeric coiled coil, which is characterized with an extended hydrophobic core with-

out “polar defects” and a large proportion of networked salt bridges. Both folding and unfolding appear to represent structural transitions between folded trimer and unfolded monomer, without detectable intermediates. From the kinetic rate constants of refolding and unfolding the free energy of unfolding was calculated as $\Delta G_u \sim 76$ to 88 kJ mol^{-1}, in good agreement with $\Delta G_u = 79 \pm 10$ kJ mol^{-1} obtained by equilibrium unfolding. Our estimates of ΔG_u is almost half of ΔG_u recently suggested from thermal unfolding experiments in the absence of denaturant. Since GdmCl attenuates electrostatic interactions, it is intuitive to ascribe the discrepancy to an electrostatic component of the unfolding barrier. We found that Lpp-56 traverses an unstructured and highly solvent-accessible transition state. We investigated the electrostatic properties of Lpp-56 by computational methods. There is an unusually large electrostatic contribution of electrostatics to the stability of Lpp-56. Screening of charge-charge interactions is in part responsible for the discrepant estimates of the unfolding free energy. The long lifetime of rings of networked salt bridges is most likely responsible for the extremely low unfolding rate (kinetic stability) by adding a serious electrostatic component to the high-energy barrier for unfolding.

The C-terminal domain of the end binding protein 1 (EB1c) is a dimeric coiled coil, which, however, is stabilized C-terminally by a four-helix bundle. Characterization of the stability and the folding/unfolding kinetics is indispensable toward understanding of the role of the EB1 protein within the protein network controlling the dynamical properties of microtubule. Calorimetric and spectroscopic experiments lead to the conclusion that intermediary, partly folded states are populated both at equilibrium and non-equilibrium conditions. The folding free energy estimates from equilibrium and kinetic experiments differ by much if urea is used as the denaturing agent. Discrepant results were obtained from kinetic experiments performed in urea and GdmCl: The apparent folding rate is higher and the apparent unfolding rate is lower in aqueous buffer if the linear extrapolation model is applied to data collected in GdmCl. The data can be tentatively rationalized in terms of electrostatic effects in the heavily negatively charged protein at neutral pH.

The SIV glycoprotein 41 (gp41) is a six-helix bundle, in which three N-terminal helices form a parallel coiled coil trimer and the C-terminal helices pack in the reverse direction in the grooves on the surface of the central coiled coil. We were interested to probe by mutational analysis the energetic and kinetic role of topologically conserved contacts which stabilize the internal coiled coil and anchor the outer helices on its surface. To our knowledge this the first study aiming at characterization of the transition state of a trimeric protein. The data demonstrate that some mutations cause large deviations from the two-state behavior. Nonetheless, nine variants obey a weakened two-state criterion. “Quasi-Φ-value” analysis and β_T-value analysis lead to the conclusion that the six-helix bundle transverses a compact folding transition state, in which native-like structure formation is modestly advanced, yet there are no completely unstructured regions. We propose that the highest-energy barrier along the folding pathway is passed in the trimeric state, after the C-terminal half of each monomer chain is fixed in anti-parallel orientation to the surface of the central N-terminal coiled-coil-to-be.

ZUSAMMENFASSUNG

Trotz des schnellen Fortschritts und wichtiger Ergebnisse in den Gebieten der Strukturbiologie und Biophysik der Proteine bleiben viele Aspekte des Proteinfaltung Problems im Dunkeln. Die fundamentale Ursache dafür besteht in der konzeptionellen Vielschichtigkeit der Fragestellung Wie faltet sich eine ungefaltete Polypeptid-Kette in eine eindeutige dreidimensionale Struktur. Proteine falten entlang einer ausgedehnten, rauen und mehrdimensionalen Energieoberfläche, die von energetischen Faktoren bestimmt ist, welche gross im Wert sind, sich aber gegenseitig kompensieren. Der Faltungsprozess beinhaltet die Bildung einer Vielzahl von nicht-kovalenten Bindungen, einer erheblichen Reduktion der Freiheitsgrade und eine massive Verkleinerung der zum Wasser exponierten Molekülоberfläche. Aufgrund der Grösse von Proteinen, ihrer strukturellen Diversität und ihres ensembleartiges Verhalten sind weder theoretische Modelle noch Experimente in der Lage die Feinheiten der Energieoberfläche und die mechanistische Abfolge der Ereignisse auf dem Weg zum gefalteten Molekül zu beschreiben.

Die Mehrheit der heutigen Erkenntnisse über Proteinfaltung entstammen von Untersuchungen kleiner globulären Proteinen oder globulärer Domänen von grösseren, mehrdomänigen Proteinen. Andererseits sind Proteine, die aus zwei oder mehr nicht-kovalent miteinander verbundenen Polypeptid-Ketten bestehen, viel häufiger in der Natur anzutreffen. Diese Oligomere entwickelten sich zu derjenigen evolutionären Lösung, die den Druck von physikalischen und chemischen Faktoren in Einklang mit der Diversifikation der biologischen Funktion bringt. Monomere und oligomere Proteine werden durch dieselben Kräfte und thermodynamischen Mechanismen stabilisiert. Trotzdem ist die Faltung von oligomeren Proteinen in zwei Aspekten aussergewöhnlich: Erstens besteht in der Folge der molekularen Ereignisse vom ungefalteten zum gefalteten Zustand zwingend ein Schritt in dem sich die Molekularität ändert. Zweitens agieren die ubiquitären energetischen Mechanismen der Proteinstabilisierung zur selben Zeit sowohl auf den Ebenen der einzelnen Untereinheiten als auch des ganzen Oligomers. Dies bedeutet, dass die molekularen Wechselwirkungen zwischen den Untereinheiten wie auch innerhalb derer sich gegeneinander ausgleichen. Als Konsequenz ergibt sich, dass oligomere Proteine grundsätzlich verschiedende Faltungswege beschreiten. Deutlich mehr Aufwand ist erforderlich, um die Faltungsdetails von oligomeren Proteinen zu erforschen, um unsere Kenntnisse über die Proteinfaltung im allgemeinen zu vertiefen und somit schliesslich deren grossen Potential für biomedizinische und biotechnologische Anwendungen auszunutzen.

Die vorliegende Dissertation widmet sich der biophysikalischen Charakterisierung von

drei oligomeren Proteinen. Alle drei teilen ein gemeinsames strukturelles Motiv eine α-helikale Coiled Coil-Struktur. Für unsere Untersuchungen wählten wir mit Absicht Proteine, die Coiled Coil-Strukturen enthalten, aber strukturelle oder biophysikalische Besonderheiten besitzen.

Der strukturelle Aufbau desäusseren Membranproteins (Lpp-56) aus E. coli besitzt eine trimere Coiled Coil-Struktur aus sieben Heptaden, die sich durch einen ausgedehnten hydrophoben Kern ohne polare Defekte und einer grossen Zahl von untereinander vernetzten Salzbrücken auszeichnet. Sowohl das Faltungs- wie auch das Entfaltungsverhalten deuten auf strukturelle bergänge zwischen gefalteten Trimeren und ungefaltenen Monomeren ohne messbare Intermediaten hin. Aus den (kinetischen) Geschwindigkeiten von Rückfaltung und Entfaltung wurde die freie Energie ΔG_u der Entfaltung von ~76 to 88 kJ mol^{-1} berechnet, welche in guter bereinstimmung mit dem Wert von $\Delta G_u = 79 \pm 10$ kJ mol^{-1} ist, der in Gleichgewicht-Faltungsexperimenten bestimmt wurde. Unser Messwert von ΔG_u ist nur halb so gross wie der kürzlich veröffentlichte Wert für ΔG_u, der mit Hilfe von thermischen Denaturierungsexperimenten in Abwesenheit eines Denaturans bestimmt wurde. Da Guanidiniumchlorid die elektrostatischen Wechselwirkungen abschwächt, ist es naheliegend, den Unterschied dem elektrostatischen Beitrag der Entfaltungsbarriere zuzuschreiben. Wir haben gezeigt, dass bei der Ent-/Rückfaltung von Lpp-56 das Protein einen unstrukturierten und hochgradig wasserexponierten bergangszustand einnimmt. Durch Computer-gestütze Methoden haben wir die elektrostatischen Eigenschaften von Lpp-56 untersucht. Dabei haben wir festgestellt, dass die Elektrostatik einen ungewöhnlich grossen Beitrag zur Stabilität von Lpp-56 beiträgt. Abschirmung von Ladung-Ladung Wechselwirkungen haben sich zum Teil als verantwortlich für die abweichende Bestimmungen der freien Energie der Entfaltung erwiesen. Die lange Lebensdauer der Salzbrücken ist höchstwahrscheinlich für die ausnehmend kleine Entfaltungsrate (kinetische Stabilität) verantwortlich, indem sie einen erheblichen elektrostatischen Beitrag zur hochenergetischen Barriere der Entfaltung beisteuert.

Die C-terminale Domäne des end binding protein 1 (EB1c) besteht aus einer dimeren Coiled Coil-Struktur, die aber durch ein Bündel aus vier Helices am C-Terminus stabilisiert wird. Die Charakterisierung der Stabilität und der Ent-/Faltungskinetik ist unerlässlich für das Verständnis der Rolle des EB1 Proteins innerhalb des Protein Netzwerkes, das die dynamischen Eigenschaften der Mikrotubulen kontrolliert. Kalorimetrische und spektroskopische Experimente führten zur Schlussfolgerung, dass intermediäre, teilweise gefaltete Zustände sowohl unter Gleichgewichts- wie auch Nicht-Gleichgewichts-Bedingungen vorhanden sind. Die freie Energie der Faltung aus Gleichgewichts- und kinetischen Experimenten unterscheiden sich stark, wenn Harnstoff als Denaturans verwendet wird. Unterschiede ergeben sich auch zwischen kinetischen Experimenten mit Harnstoff und Guanidiniumchlorid: Unter Annahme des Linearen-Extrapolations-Modells ist die apparente Faltungsgeschwindigkeit in wässriger Lösung mit GdmCl-Denaturierung höher und die apparente Entfaltungsgeschwindigkeit niedriger. Eine mögliche Erklärung des oben beschriebenen Verhaltens kann die starke negative Ladung des Proteins bei neutralem pH-Wert sein.

Das SIV Glykoprotein 41 (gp41) besteht aus einem Bündel aus sechs Helices, in dem drei N-terminale Helices ein parallel ausgerichtetes Coiled Coil-Trimer bilden und die C-terminalen Helices sich in antiparalleler Richtung in die Gruben der zentralen Coiled Coil einnisten. Unsere Untersuchungen beabsichtigten durch Analyse von mutierten Proteinen den energetischen und kinetischen Beitrag von konservierten topologischen Kontakten, die die zentrale Coiled Coil-Struktur stabilisieren und die äusseren Helices auf deren Oberfläche verfestigen, zu bestimmen. Unseres Wissens handelt es sich hierbei um die erste Untersuchung mit dem Ziel den bergangszustand eines trimeren Proteins zu charakterisieren. Unsere Resultate zeigen, dass einige Mutationen eine grosse Abweichung von einem Zwei-Zustand Modell zeigen. Trotzdem gehorchen neun Varianten dem abgeschwächten Zwei-Zustand Modell. Die Analyse der "Quasi-Φ-Wertre" und β_{T}-Werte lässt die Schlussfolgerung zu, dass bei der Faltung das Sechs-Helix-Bündel einen kompakten, der nativen Struktur ähnlichen bergangszustand durchläuft, der keine vollständig unstrukturierten Bereiche mehr aufweist. Unserer provisorischen Vorstellung nach, wird die Barriere höchster Energie im Lauf des Faltungsweges in einem trimeren Zustand überschritten, nachdem die C-terminalen, momeren Fast-Helices sich in antiparalleler Orientierung an die Oberfläche der naszenten, zentralen, N-terminalen "Coiled Coil"-Struktur angelagert haben.

CHAPTER 1

INTRODUCTION

Proteins are polymers made of 20 ”standard” amino acids. They comprise the main part of the cell machinery by fulfilling structural, metabolic and regulating functions. The chief characteristic of proteins is their ability to bind and modify other molecules such as DNA, other proteins or small molecules. Although recent studies point to the presence of disordered yet functional polypeptide chains in the cell, in most cases biologically active proteins exhibit a unique three dimensional structure. After the pioneering work of Anfinsen [1] it has become clear that relatively short polypeptide chains (50-300 amino acids) are capable to fold spontaneously into the biologically relevant conformation. In the following discussions I will consider only autonomously folding protein domains since they are amenable to biophysical characterization. The chaperon-assisted folding of larger proteins is beyond the scope of this work.

In the test tube, folding can be described as a dynamic equilibrium reaction between a huge ensemble of disordered polypeptide chains being close in energy, the unfolded state, and the native state, which is also an ensemble but is more localized in the conformational space. At benign or physiological conditions the native state is more populated. Interestingly, the native state is only marginally more stable than the unfolded state - rarely by more than 40-50 kJ mol^{-1}. Given that the cumulative magnitude of non-covalent interactions that stabilize the 3D structure is on the order of thousands of kJ mol^{-1} it is clear that proteins fold as the result of a very precise balance between opposing physical forces and thermodynamic effects. The water plays an essential role in tuning this balance. It is still not completely understood to what extent have biological factors directed the evolution of proteins towards marginal stability. However, it appears plausible that proteins must be stable enough to resist destabilizing, random mutations and variation of the environmental conditions. On the other hand proteins that are very stable tend to be also rigid, while many biological tasks require some structural plasticity.

Sufficiently stable species intermediary in structure and energy between the folded and unfolded states are only rarely populated at benign conditions. In contrast, the protein passes through a series of transient intermediates in the process of folding (and unfolding). Such intermediates typically guide the folding process: In a folding pathway perspective (or folding funnel perspective) they can be simplistically regarded as buoys with increasing amount of native-like structure and decreasing solvent exposure, the closer they are to the native state. From an energetic perspective, such intermediates are not populated at equilibrium since they exhibit poor optimization of enthalpic (packing) interactions in view of the large entropic losses due to the collapse of the polypeptide chain and the incomplete dehydration of hydrophobic side chains. Proteins can fold via parallel or branched pathways [2][1]. In any case, as every chemical reaction, protein folding passes through a very high-energy transition state (TS). The height of the energetic barrier separating the unfolded state and TS determines the rate of the folding reaction. There have been many efforts to understand why proteins fold on time scales ranging from microseconds to seconds [3]. One important task in this endeavor is to elucidate the structure of TS. Unfortunately, TS cannot be seen by structural methods. All existing approaches are based on indirect experimental evidence [4], or are based on computation [5].

Despite the significant advances that have been made over the past decades our current understanding of the energetics, kinetics and the structural determinants of protein folding is insufficient. This is clear from the persistent difficulties to design artificial proteins from scratch. Most of our knowledge on the biophysics of proteins has been obtained by study of monomeric proteins. Interestingly, according to current estimates oligomeric proteins are probably more abundant in nature. It appears that there is strong evolutionary pressure for monomeric proteins (or individual polypeptide chains) to associate into oligomers .

1.1 THE EVOLUTIONARY BENEFIT OF OLIGOMERIC PROTEINS

Biologically active are both short polypeptides (dipeptides even) and extremely long polypeptide chains ($<$ 20000 amino acids). However, more than half of the known proteins have molecular size between 100 and 300 amino acids (25 to 50 kDa). Why is a larger, but not too large protein better coping with the available physical-chemical-functional space available for life? Proteins (in most cases) exert their function being uniquely structured, and the realization of the 3D structure is the result of oppositely directed physical forces - folding is an enthalpically favorable process but is linked to significant entropic penalty. Only formation of a multitude of strong enthalpic interactions in a large hydrophobic core can over-compensate the entopic expenditures linked to the loss of conformational freedom of the polypeptide chain. Larger proteins bury more effectively non-polar surface from water. A larger globule of amino acids reduces the osmotic pressure since it partially buries charges being otherwise bound to neutralizing ions and water molecules. The larger the surface of a protein, the more effective the binding of ligands is, since

---

[1] Kinetic intermediates can also be off-pathway. Such chain configurations are relatively stable but involve large amounts of non-native like structure. Folding is slowed down because time elapses until such intermediates unfold or rearrange and enter a productive pathway.

the ligand is attracted by non-specific forces and is allowed to find the specific binding site by two-dimensional diffusion [6]. Also, the need for creating morphological structures and multitasking cellular machines has driven the increase in length/size of biological polypeptides. Cellular machines mentioned, there has been an evolutionary pressure for combining complementary functions within a single protein, leading to the appearance of multi-domain proteins. The advantage of domain organization is out of question. Domain organization had led to the appearance of cooperativity (via subtle conformational rearrangements), substrate/product feedback, and substrate tunneling in enzyme catalysis. Furthermore, a multi-domain protein confers specificity in the context of different subunit interactions. Indeed, catalytic sites are often found at the interface of sub-units. From a biophysical perspective, finding of the biologically relevant conformation could be faster if separate segments of the polypeptide chain fold autonomously.

It appears that there have been two parallel evolutionary pathways leading to multi-domain proteins. Large proteins typically exhibit domain organization. However, given the constant probability of translation errors, long polypeptide chains are subjected to a higher probability for a deleterious error. Oligomeric proteins built up from non-covalently associated sub-units have provided the solution. Synthetic errors are reduced for smaller polypeptides. For reasons of molecular economy, minimization of the size of structural genes and reducing the costs for elimination of misfolded polypeptide chains, oligomers represent a valuable alternative. Oligomeric proteins have opened the possibility for a more stringent selection, since mutational effects are amplified. Point mutations simultaneously affect all n-monomers comprising a homo-n-meric protein. In terms of function, next to recognition of symmetric binding sites, oligomers facilitate the existence of powerful controlling mechanisms via concentration-dependent shifts in the population of different gene products. Not surprisingly, there has been strong evolutionary pressure for monomeric proteins to associate into oligomers. According to some estimates, oligomers are more abundant in the cell.

1.2 Stability and folding of oligomeric proteins

As already mentioned, the present-day understanding of protein stability and protein folding is mostly gained from studies of monomeric proteins. Nevertheless, many of the obtained results can be also applied to oligomeric proteins. The stability of proteins, regardless if their oligomeric state is result of opposite forces. The hydrophobic effect (van der Waals interactions between non-polar groups in the folded protein and the concomitant favorable entropy of desolvation of non-polar surface), hydrogen bonds, and attractive electrostatic interactions between charged residues stabilize the native state. They are opposed by the loss of conformational degrees of freedom upon folding . Energetically, there is no principal difference between folding mechanism of monomeric and oligomeric proteins. Also, both monomers and oligomers can exhibit simple one-step folding pathway, or fold by passing through intermediary, partly folded states. However, one difference is important. At least one step in the folding pathway of oligomeric proteins must be concentration dependent. The reason is that the formation of the ologomeric folded structure requires folding and association. This means that the population of the biologically active oligomer and the rate of folding are concentration-dependent. The nature has used this pecularity to create powerful regulation mechanisms by controlling experession levels.

In some cases folding and oligomerization can be described as separate and virtually independent processes (for example cAMP receptor protein, disulfoferredoxin , dyneine light chain). On the other hand there are cases where folding and oligomerization are fully coupled processes (for example Arc repressor [7], GCN4 leuzine zipper [8], mannose-binding lectin [9]). The degree of coupling between folding and association can be described as balance between intra- and intermolecular forces that stabilize the ologomeric structure. If the monomeric subunits are stabilized by intermolecular forces resulting in thermodynamically stable structures, then folding and association can be considered as separate steps: the formation of oligomers corresponds to the association of prefolded monomers. In contrast, when the monomeric subunits are unstable and stability is gained predominately by intermolecular forces within the ologomeric structure, then folding and association closely depend on each other. The coupling between the concurrent processes of folding and association, and the dependence of folding on the protein concentration adds an extra level of complexity. This is perhaps the mainly reason that our knowledge on the stability and folding mechanisms of oligomers is limited to only a few well characterized proteins. In the following I briefly sketch possible folding pathways of oligomeric proteins.

Folding of dimeric proteins can follow the simple two state model (2SM), from the unfolded monomers to the folded dimers [7]. In more complicated cases the monomers are partially prefolded and associate to the folded dimer [8]. In even more complicated cases, next to partly folded monomers, partially folded oligomeric species can be observed in the folding pathway. For statistical reason any reaction higher then second order is very slow [2]. So it may be surprising that there are trimeric proteins, for which folding according to the 2SM has been reported [10, 11]. Is it possible that that three unfolded monomers fold to the fully folded trimer without any intermediates? This scenario can occur if the dimer is very unstable and sparsely populated, so that the equilibrium between monomer and dimer is on the side of the unfolded monomer, and simultaneously, the equilibrium between dimer and trimer is far on the site of folded trimer. Under these conditions, the trimer formation appears as two-state reaction. It must also be considered that the dimeric intermediate can be “silent” depending on the experimental techniques used. Indeed, the applicability of the 2SM to experimental data does not implicate that there are no intermediates at all. Intermediates are always present in the folding of a protein, but they are either not populated (kinetically instable) or cannot be detected by the available experimental methods for reasons of sensitivity and/or time resolution. So far, simple one-step folding of a tetrameric protein has not be reported. Tetramers can fold only through intermediates. Indeed, the folding of the teramerization domain of p53 tumor suppressor and bacterial R67 dihydrofolate reductase passes over dimeric intermediates and can be formally described as two consecutive bimolecular reactions.

In summary, the native state of oligomeric proteins can be achieved in different ways, depending on the balance of inter- and intramolecular forces. Perhaps the most enigmatic case are proteins in which folding and association are inseparable events. Coiled coils represent the archetypal example of this type of folding. This thesis is devoted to biophysical characterization of coiled coil proteins.

1.3 Structural properties coiled coils

The coiled coil motif is a ubiquitous protein structure motif. It is found in more then 7% of all naturally occurring proteins [12]. Coiled coils participate in different types of biologically important processes: muscle contraction, transcription, metabolism, transmembrane channeling, chaperon-assisted folding, etc. In 1953 Crick [13] proposed that the main structural element of a large class of fibrous proteins is represented by two or more α-helices wrapped around each other to form a superhelix, as in keratin, myosin and fibrinogen. The central characteristic of these proteins is a sequential repeat of seven amino acids, named heptad, in which hydrophobic side chains show the characteristic 4, 3 repetive pattern (see below). The definitive proof of Cricks proposition has come almost thirty years later, in 1981, when the crystal structure of the influenza virus hemagglutinin was solved [14]. This was the first high-resolution view if a three-stranded coiled coil. Further seven years were needed to obtain an atomic-resolution structure if a simple dimeric coiled coil - the leucine zipper subdomain of the yeast transcription factor GCN4 [15].

A coiled coil is a bundle of α-helices that are wound into a superhelix. There may be two, three, four or more helices in the bundle, and they may run either in the same (parallel) or in opposite (anti-parallel) direction. The characteristic of coiled coils is the distinctive packing of amino acid side chains in the core of the bundle, called "knobs-into-holes". In this type of packing one residue (knob) fills a hole made of four residues of the adjacent helix. The sequential heptad repeat contains seven consecutive amino acids which are denoted as **a-b-c-d-e-f-g**. In a typical coiled coil, positions **a** and **d** are occupied by non-polar side chains and form the inrerhelix interface, which comprises the hydrophobic core. The other heptad positions are hydrophilic and form the solvent exposed part of the coiled coil.

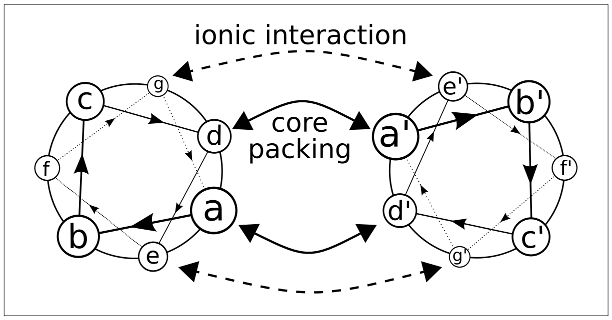


Fig. 1.1: Coil coild interaction patern in a dimeric protein. The typical interaction pattern of hydrophobic residues in positions **a** and **d** and ionic interation at positions **e** and **g** are indicated by the solid and dashed double-headed arrows, respectively.

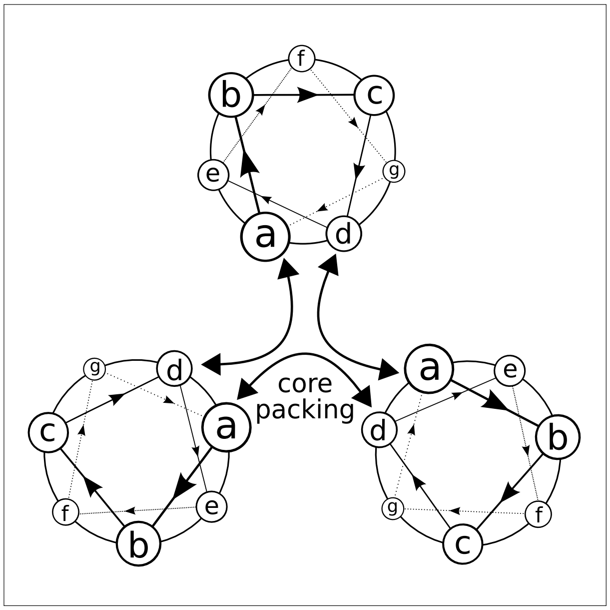


Fig. 1.2: Helix-wheel representation of a trimeric coiled coil.

The strandneness, i.e. the number of α-helices building up a coiled coil is mainly dictated by the amino acids occupying the **a** and **d** positions. This is because the residues in the **a** and **d** positions in the two-, three-, four-, five-, six-stranded coiled coils have different geometry in respect to the backbone of the adjacent helix, resulting in different side chain preferences. For example, isoluecine in the **a** position in combination with leucine in the **d** position drives formation of dimeric coiled coils. If isoleucine, leucine and valin are distributed over the **a** and **d** positions the coiled coil is preferably trimeric. Leucine in the **a** positions and isoleucine in the **d** positions tend to promote tetramerization (see [12] and references therein). The core-flanking, solvent exposed positions **e** and **g** mainly influence the orientation of the α-helices. As these residues are usually polar, and mostly charged, electrostatic repulsion/attraction contributes critically to formation of parallel or anti-parallel coiled coils [16]. To achieve the regular meshing of contacts at the interhelix interface the side chains occupying identical heptad positions must have identical geometrical orientation in respect to the α-helix axis, turn after turn. This is not possible with regular α-helices, which have approximately 3.6 residues per turn, leading to a continous drift of the side chain position. In the coiled coils the problem is resolved by a slight distortion of the canonical α-helices geometry (left handed twist) so that the there are 3.5 residues per turn, thus allowing identical geometrical orientation of the side chains of every seventh position. Figs. 1.1 and 1.2 illustrates the typical pattern of molecular interactions in dimeric and trimeric coiled coils.

1.4 Objectives of the present work

A survey of the literature shows that biophysical studies on the stability, dynamics and folding of coiled coils focus almost exclusively on short dimeric motifs, more than three quarters of the published work being done with the GCN4 leucine zipper and derivatives thereof. The biophysical properties of trimers and higher-order coiled coils are much less well known. For our investigations we on purpose selected proteins which contain coiled coils but exhibit structural or biophysical peculiarities. Details on the proteins are given further below in the corresponding chapters. Here, I mention only the main features of these proteins, which make them an interesting experimental object and motivated the presented studies.

- The protein moiety of the *E. coli* outer membrane protein (Lpp-56) is a seven heptades long trimeric coiled coil, which is characterized with an extended hydrophobic core without polar defects and a large proportion of networked salt bridges [17]. Preliminary work from this laboratory and results by others have demonstrated that the thermal unfolding transition of the protein is apparently kinetically irreversible (on the time scale accessible for experimentation), yet is thermodynamically completely reversible. We ask whether the stability of the protein can be determined by direct equilibrium experiment(s) and whether the extremely slow approach to equilibrium is dominated by slow refolding rate, or slow unfolding rate. We further were interested to characterize the role of the seemingly intensive electrostatic interactions to the stability of the trimer.

- The C-terminal domain of the end binding protein 1 (EB1c) is a dimeric coiled coil, which, however, is stabilized C-terminally by a four-helix bundle [18]. This protein exhibits unusually high ratio of acidic-versus-basic groups (theoretical pI $<$ 4). There is evidence that the three highly conserved isoforms of the EB protein can form in vivo homodimers and heterodimers. Therefore, characterization of the stability and the folding/unfolding kinetics is indispensable toward understanding of the biology of the protein. As a first step, we performed equilibrium and kinetic folding/unfolding experiments with the naturally most abundant isoform 1, and collected some initial data on the thermal stability of isoforms 2 and 3. Comparison of equilibrium and kinetic data from experiments using guanidinium chloride or urea as denaturants was intended as one possible approach to understand the consequences of the charge disbalance.

- The SIV glycoprotein 41 (gp41) is a six-helix bundle, in which three N-terminal helices form a parallel coiled coil trimer and the C-terminal helices pack in the reverse direction in the grooves on the surface of the central coiled coil [10, 19, 20]. We were interested to probe by mutational analysis the energetic and kinetic role of topologically conserved contacts which stabilize the internal coiled coil and anchor the outer helices on its surface.

BIBLIOGRAPHY

[1] Anfinsen, C. B & Scheraga, H. A. (1975) Experimental and theoretical aspects of protein folding. *Adv. Protein Chem.* **29**, 205–300.

[2] Buchner, J & Kiefhaber, T, eds. (2005) Protein Folding Handbook. (Wiley-VCH).

[3] Fersht, A. (1998) Structure and Mechanism in Protein. Sci.ence: A Guide to Enzyme Catalysis and Protein Folding. (Freeman, W. H., Publishers).

[4] Wendt, H, Leder, L, Härmä, H, Jelesarov, I, Baici, A, & Bosshard, H. R. (1997) Very rapid, ionic strength-dependent association and folding of a heterodimeric leucine zipper. *Biochemistry* **36**, 204–213.

[5] Fersht, A. R & Daggett, V. (2002) Protein folding and unfolding at atomic resolution. *Cell* **108**, 573–582.

[6] Dürr, E, Jelesarov, I, & Bosshard, H. R. (1999) Extremely fast folding of a very stable leucine zipper with a strengthened hydrophobic core and lacking electrostatic interactions between helices. *Biochemistry* **38**, 870–880.

[7] Srivastava, A. K & Sauer, R. T. (2000) Evidence for partial secondary structure formation in the transition state for arc repressor refolding and dimerization. *Biochemistry* **39**, 8308–8314.

[8] Dragan, A. I & Privalov, P. L. (2002) Unfolding of a leucine zipper is not a simple two-state transition. *J. Mol. Biol.* **321**, 891–908.

[9] Bachhawat, K, Kapoor, M, Dam, T. K, & Surolia, A. (2001) The reversible two-state unfolding of a monocot mannose-binding lectin from garlic bulbs reveals the dominant role of the dimeric interface in its stabilization. *Biochemistry* **40**, 7291–7300.

[10] Chan, D. C, Fass, D, Berger, J. M, & Kim, P. S. (1997) Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **89**, 263–273.

[11] Marti, D. N, Bjelić, S, Lu, M, Bosshard, H. R, & Jelesarov, I. (2004) Fast folding of the HIV-1 and SIV gp41 six-helix bundles. *J. Mol. Biol.* **336**, 1–8.

[12] Lupas, A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.* **21**, 375–382.

[13] Watson, J. D & Crick, F. H. (1953) The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18**, 123–131.

[14] Wilson, I. A, Skehel, J. J, & Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. *Nature* **289**, 366–373.

[15] Wendt, H, Drr, E, Thomas, R. M, Przybylski, M, & Bosshard, H. R. (1995) Characterization of leucine zipper complexes by electrospray ionization mass spectrometry. *Protein. Sci.* **4**, 1563–1570.

[16] Oakley, M. G & Kim, P. S. (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* **37**, 12603–12610.

[17] Shu, W, Liu, J, Ji, H, & Lu, M. (2000) Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 A resolution. *J. Mol. Biol.* **299**, 1101–1112.

[18] Hayashi, I & Ikura, M. (2003) Crystal structure of the amino-terminal microtubule-binding domain of end-binding protein 1 (EB1). *J. Biol. Chem.* **278**, 36430–36434.

[19] Tan, K, Liu, J, Wang, J, Shen, S, & Lu, M. (1997) Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl. Acad. Sci. U. S .A.* **94**, 12303–12308.

[20] Weissenhorn, W, Dessen, A, Harrison, S. C, Skehel, J. J, & Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426–430.

CHAPTER 2

STABILITY AND FOLDING OF THE *E. coli* OUTER MEMBRANE LIPOPROTEIN - THE ROLE OF ELECTROSTATIC INTERACTIONS.

2.1 INTRODUCTION

The E. coli outer membrane lipoprotein, Lpp, is the most abundant protein in the *E. coli* cell. It is a very important component in the organization of the bacterial cell wall (see Fig. 2.1A). The 56 amino acids long protein moiety of Lpp, named Lpp-56, is a soluble parallel trimeric coiled coil [Shu et al., J. Mol. Biol. 299: 1101-1112] and Fig. 2.1C. Details on the biology and the structural peculiarities of the molecule are given in the following articles. Here I briefly describe the experimental information, which motivated the presented studies.

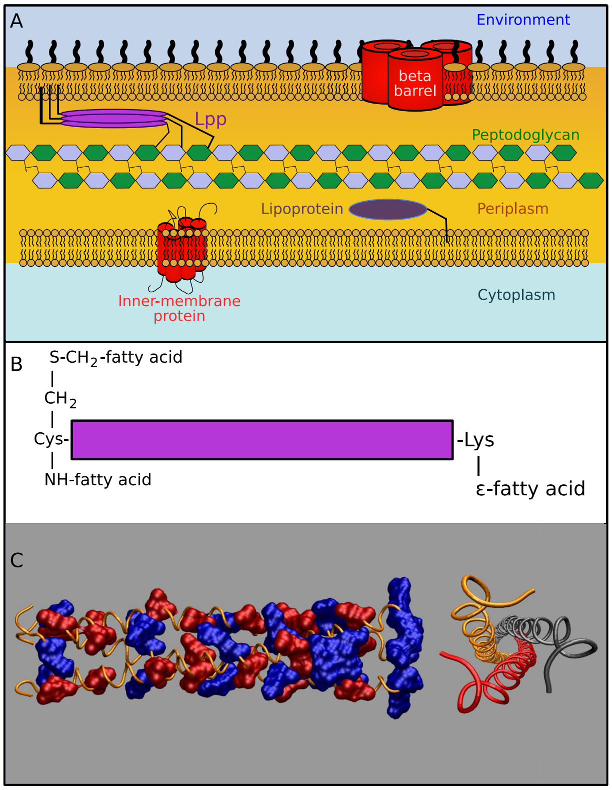


Fig. 2.1: Molecular organisation, funcion and structure of Lpp. (Panel A) The localization of the Lpp protein in the bacterial cell wall is presented. As shown in the scheme on panel B, the protein moiety is esterified and thioesterified to fatty acids via the N-terminal cysteine and is covalently bound to the peptidoglycan matrix via the ϵ-amino group of the C-terminal lysine. In this way it becomes an integral part of the cell wall. Panel C, side and top representation of of the Lpp-56. In the side representation the helces are represented as tube and cahrged residues as surface, whereas glutamic/aspartic acid and arginine/lysine are in red and blue, respectively. The picture in panel C was created using the atomic coordinates as deposed in pdb entry 1eq7.

CD and DSC experiments demonstrate that the thermal melting behavior of Lpp-56 is peculiar. The cooling traces after heating to temperatures where the protein is unfolded show pronounce hysteresis, as exemplified with CD data in Fig. 2.2A. The shape of conformational transition is completely different in heating and cooling experiments. After cooling both the ellipticity and the heat capacity of the sample differ by much from the initial values.Furthermore, the apparent melting temperature (estimated as the temperature where the change in ellipicity is half of the maximal, or where the heat capacity function peaks) significantly increases upon increasing the slope of the temperature

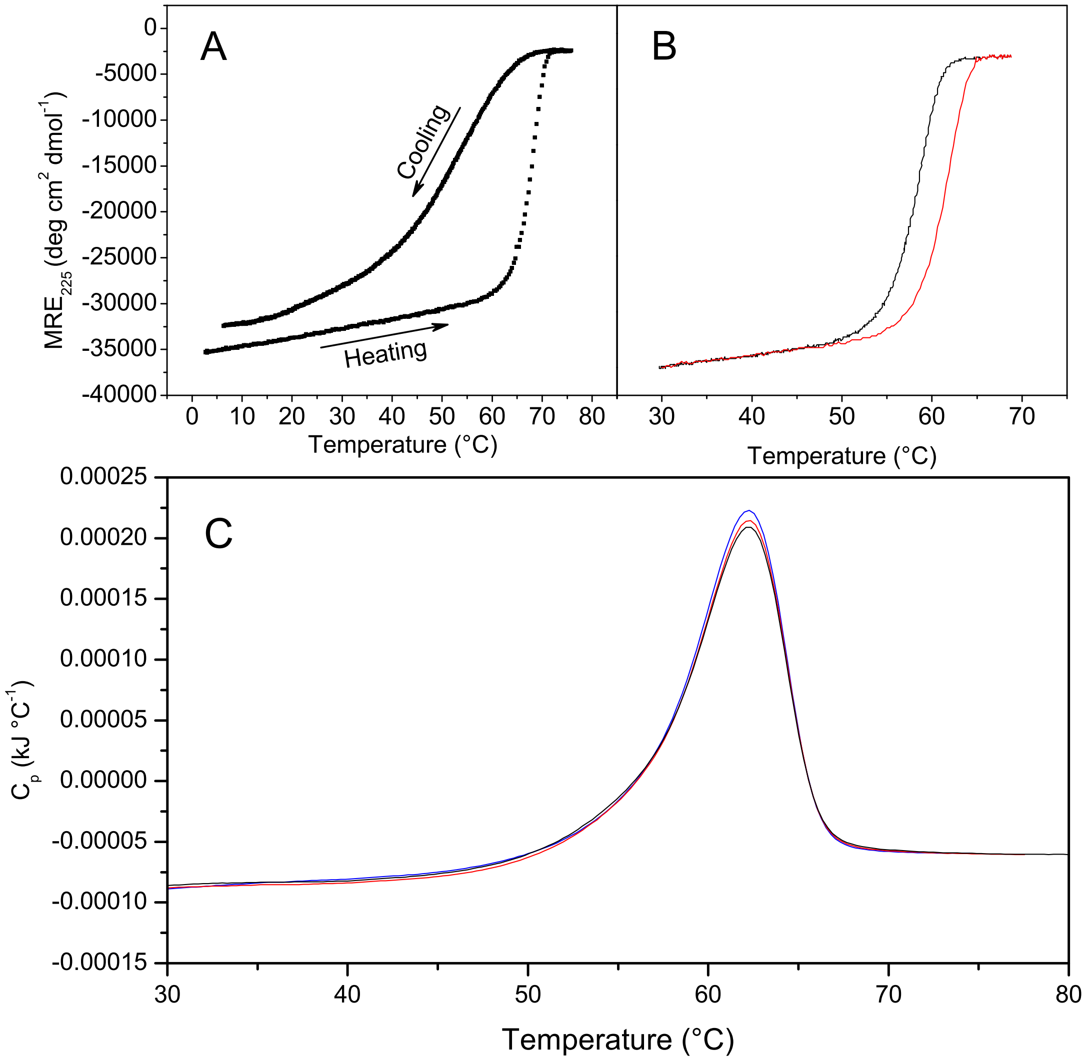


Fig. 2.2: Thermal unfolding of wild type Lpp-56 and Ala-to-Gln mutants monitored by far-UV CD. (Panel A) Thermal melting exhibits pronounced hysteresis, as exemplified for 100 M Lpp-56 at a scanning rate of 1 ℃ min^{-1}. (Panel B) Melting curves are shifted to higher temperatures when heating is faster. 100 μM wild type Lpp-56 was heated at a rate of 0.2 ℃ min^{-1} (black trace) or 1 ℃ min^{-1} (red trace).)Panel C) Temperature dependence of the partial molar heat capacity measured by DSC. Overlayed are four consecutive scans of the same sample after rapid cooling and incubation at 5 ℃ for ∼ 4 hours. Experiments were performed in cocktail buffer, pH 5 (phosphoric, citric and boric acid, 7.5 mM each, adjusted to pH 5 with KOH and to ionic strength of 0.1 M with KCl) ramp rate (Fig. 2.2B). Both observations are typical for irreversible thermal denaturation. On the other hand, the extent of hysteresis, i.e. the difference between the heating and cooling traces decreases upon increasing the protein concentration. Also, prolonged incubation of the "irreversibly denatured" protein at low temperatures leads to regain of the initial signal and the following heating trace overlays the previous. As illustrated in Fig. 2.2C, up to four DSC experiments performed in this way with the same sample yield practically indistinguishable results (the calorimetric enthalpies measured in scans 2, 3, and 4 represent > 97 % of the enthalpy of scan 1). Altogether, these observations can be interpreted to reflect an apparent kinetic irreversibility of the folding/unfolding transition, which is genuinely reversible in thermodynamic sense. This means that the rates of folding and unfolding, and their temperature dependencies are balanced in such a way that the system cannot reach equilibrium (and is actually quite far from it) on the time scale of the typical experiment (scanning rate of 0.2-1.0 ℃ min^{-1}; for technical reasons heating and cooling at scanning rates < 0.2 ℃ min^{-1} is impossible). We ask the question whether the thermodynamic stability of this interesting trimeric coiled coil is accessible from denaturant-induced isothermal experiments, and whether direct determination of the folding and unfolding rates by Chevron plot analysis is possible. The results and their interpretation are discussed in Section 2.2. The main conclusion is that there is a large electrostatic component both to the equilibrium unfolding free energy change and to the height of the unfolding energy barrier. In Section 2.3 a computational analysis of the electrostatic properties of Lpp-56 is presented.

2.2 Stability and folding/unfolding kinetics of the homotrimeric coiled coil Lpp-56

Saša Bjelić and Andrey Karshikoff and Ilian Jelesarov
Article published in Biochemistry 45: 8931-8939

Biochemistry **2006,** *45,* 8931−8939 8931

Stability and Folding/Unfolding Kinetics of the Homotrimeric Coiled Coil Lpp-56†

Saša Bjelić,‡ Andrey Karshikoff,§ and Ilian Jelesarov*,‡

Biochemisches Institut der Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland, and
Department of Biosciences and Nutrition, Karolinska Institute, Stockholm, Sweden

Received April 26, 2006; Revised Manuscript Received May 29, 2006

ABSTRACT: The 56 amino acid long protein moiety of the *E. coli* outer membrane lipoprotein (Lpp-56) contains a 4,3 hydrophobic heptad repeat and forms a parallel, in-register trimeric coiled coil in solution. The unconventional structural properties of Lpp-56 make it an interesting experimental object to study the folding of a trimeric coiled coil. Folding is unusually slow at low temperatures, and the rates of folding and unfolding are balanced in such a way that the thermodynamic equilibrium is established after considerable time at high temperatures or in the presence of denaturants. Here, we examine the stability and the folding/unfolding kinetics of Lpp-56 at neutral pH using GdmCl as the denaturant. Folding and unfolding appear to represent structural transitions between the folded trimer and the unfolded monomer, without detectable intermediates. For the first time, we estimated the unfolding free energy from a direct measurement at equilibrium. Our estimate of $\Delta G_U = 79 \pm 10$ kJ mol^{-1} compares very well with ΔG_U $\sim$76−88 kJ mol^{-1} obtained from the kinetic rate constants of refolding (7×10^5 M^{-2} s^{-2}) and unfolding (10^{-9} to 10^{-11} s^{-1}) and is almost half of the ΔG_U recently suggested using a different methodology (Dragan, A. I., Potekhin, S., Sivolob, A., Lu, M., and Privalov, P. L. (2004) *Biochemistry 43, 14891−14900*). Because GdmCl attenuates electrostatic interactions, the discrepancy can be partly explained by an electrostatic component of the unfolding barrier. The combined information illustrates the difficulties in obtaining a precise biophysical description of proteins that exhibit unusual kinetic properties. Lpp-56 is the first coiled coil for which a high unfolding kinetic barrier has been experimentally demonstrated.

Our current knowledge about the energetics and kinetics of protein folding originates mainly from studies of monomeric protein domains. However, oligomeric proteins are in fact more abundant in nature, as the consequence of a strong evolutionary pressure for monomers to associate into oligomers (*1*). Folding of monomers and oligomers is driven by the same type of noncovalent interactions and thermodynamic principles. In oligomers, the forces stabilizing and destabilizing the natively folded structure operate not only within the single subunit but also between the subunits. Furthermore, the formation of an oligomeric protein includes at least one concentration-dependent step because folding and association are often tightly coupled processes. If the subunits are completely (or almost completely) unstructured in isolation, the 3D structure is maintained through intersubunit interactions. The thermodynamics and folding mechanism of relatively short, dimeric α-helical coiled coils being archetypal for this class of oligomers have been studied in considerable detail (*2*). Much less is known about the energetics and kinetics of trimeric and higher order oligomeric coiled coils.

The Braun's lipoprotein (Lpp) is the most abundant protein in the *E. coli* outer membrane. It exists in a free form and in a bound form, which is covalently linked to the peptidoglycan matrix. Lpp is a major player in maintaining the architecture and function of the outer membrane. The 56 amino acid long protein moiety of Lpp (Lpp-56[1]) contains a 4,3 hydrophobic heptad repeat and forms a parallel, in-register trimeric coiled coil in solution (*3*). The crystal structure of Lpp-56 exhibits intriguing peculiarities: (i) an intricate, water-mediated hydrogen-bonding network firmly anchors an umbrella-shaped capping structure at the *N*-terminal. (ii) The three *C*-terminal tyrosine rings form well-defined van der Waals contacts with 3-fold symmetry. Both structural motifs appear to fix and stabilize the termini of the coiled coil, which are flexible and fraying in many other coiled coils. (iii) Three successive *a* and *d* layers include alanine residues. (iv) *β*-Branched residues (Val, Ile) predominantly occupy the *a* positions, and nonbranched residues (Leu, Ala, Met) exclusively occupy the *d* positions, in contrast to all other trimeric coiled coils known. (v) Finally, the supercoil radius of the alanine-zipper region is 5.2 Å instead of 6.7 Å in a regular three-stranded coiled coil. As a result, Lpp-56 exhibits unique α-helical curvature. The unconventional structural properties of Lpp-56 make it an interesting experimental object to study the folding of a trimeric coiled coil.

---

† This work was supported in part by the Swiss National Science Foundation (grant 31-100197/1).

* To whom correspondence should be addressed. Phone: ++41 44 635 5547. Fax: ++41 44 635 6805. E-mail: iljel@bioc.unizh.ch.

‡ Biochemisches Institut der Universität Zürich.

§ Karolinska Institute.

[1] Abbreviations: ΔC_p, heat capacity change; f_U, molar fraction of unfolded protein; $\Delta G_U^{H_2O}$, unfolding free energy; k_f, refolding rate constant; k_u, unfolding rate constant; Lpp-56, the protein moiety of the outer membrane lipoprotein of *E. coli*; LEM, linear extrapolation method; m_{eq}, the first derivative of the observed unfolding free energy with respect to the denaturant concentration; m_f and m_u, the first derivatives of the logarithm of the folding rate constant and unfolding rate constant, respectively, with respect to the denaturant concentration; MRE$_{225}$, mean residue ellipticity at 225 nm.

10.1021/bi0608156 CCC: $33.50

A couple of years ago, we initiated a thermodynamic and kinetic investigation of Lpp-56 at pH 5, where thermal unfolding is completely reversible (see Supporting Information for some results of this study). We realized that folding is unusually slow at low temperatures, and the rates of folding and unfolding are balanced in such a way that the equilibrium between the folded trimer and the unfolded monomer is established after considerable time at high temperatures or in the presence of denaturants. As it sometimes happens in research, while our studies were in an advanced stage, Dragan et al. published an extensive study on the biophysics of Lpp-56-folding at neutral pH (*4*). (Henceforth, we refer to this work as DP, acknowledging the first and the principal authors). They presented a wealth of data documenting well the kinetic peculiarity of the protein. DP used an innovative and elegant method to derive information on the energetics of the folded state from experiments performed under nonequilibrium conditions and provided insights into the energetics of the folding transition state. One important conclusion of this study is that the free energy of unfolding at 4 °C amounts to 137 kJ mol^{-1}, corresponding to an unfolding equilibrium constant of 1.5×10^{-26} M^2. Because the directly measured rate constant of folding at 5 °C was 2.3×10^6 M^{-2} s^{-1} and both folding and unfolding could be modeled with simple $3M \rightarrow T$ and $T \rightarrow 3M$ reactions, respectively, the rate constant of unfolding should be on the order of 10^{-20} s^{-1} at benign conditions. It follows that the expected half time for Lpp-56 unfolding is on the order of 10^{12} years!

Intrigued by this extremely high kinetic stability, here we reexamine the stability and the folding/unfolding kinetics of Lpp-56 at neutral pH using GdmCl as the denaturant. For the first time, we estimated the unfolding free energy from a direct measurement at equilibrium. Our estimate of $\Delta G_U = 79 \pm 10$ kJ mol^{-1} compares very well with the $\Delta G_U \sim 76-88$ kJ mol^{-1} obtained from the kinetic rate constants of refolding and unfolding and is almost half of the ΔG_U suggested by DP using a different methodology. The discrepancy can be partly explained by an electrostatic component of the unfolding barrier. The combined information illustrates the difficulties in obtaining a precise biophysical description of proteins that exhibit unusual kinetic properties.

EXPERIMENTAL PROCEDURES

The Lpp-56 protein was provided by Dr. Min Lu from Cornell University. The expression and purification procedures were described (*3*). The sequence of the protein is given in Figure S2 of Supporting Information. The purity of the protein was checked by RP-HPLC, and the mass was verified by mass spectroscopy. Protein monomer concentration was determined by measuring UV absorbance at 280 nm in 6 M GdmCl using an extinction coefficient $\epsilon = 1280$ M^{-1} cm^{-1} calculated from the amino acid sequence (*5*). Concentrated stock solutions were prepared by dissolving the lyophylized protein either in an aqueous buffer or in a buffer containing 6 M GdmCl and by performing extensive dialysis for at least 20 h.

Buffer. All experiments were conducted in a standard PBS buffer (8.3 mM $\mathrm{Na_2HPO_4}$, 1.47 mM $\mathrm{KH_2PO_4}$, 137 mM NaCl, 2.7 mM KCl; I = 164 mM) at pH 7.0. The pH of buffers containing GdmCl was adjusted after adding the denaturant. The GdmCl concentration of each sample was verified by measuring the refractive index.

Circular Dichroism (CD) Spectroscopy. Equilibrium unfolding experiments were performed on a Jasco-715 instrument (Jasco Ltd.), the calibration of which was routinely checked. In addition, for manual mixing kinetic experiments, models J-500 and J-810 were used. All spectrometers were equipped with a computer-controlled water bath. Cylindrical jacketed cuvettes of 0.1 and 1.0 cm spectral paths were used.

Equilibrium Unfolding Experiment. The final protein concentration concentration of all samples was 12 μM (monomer equivalents) in PBS containing 0−6 M GdmCl. The protein was diluted from stock solutions prepared in either PBS or PBS containing 6 M GdmCl. Each solution was split into aliquots and stored at 4 °C in sterile tubes to prevent contamination during the experiment. Extreme care was taken not to disturb the thermal equilibrium during the measurement. The solutions were transferred using chilled pipet tips directly into the cuvette, which had been mounted in the cell-holder and extensively prethermostated. The ellipticity at 225 nm was sampled for 3 min after complete thermal equilibration and averaged over 60 data points. As a test for possible instrumental drift samples in which the equilibrium was established after a few days of incubation, were regularly remeasured. No systematic instrumental drift was detected.

Manual Mixing Kinetics. Depending on the protein concentration and GdmCl concentration, the apparent half times of folding and unfolding of Lpp-56 range from seconds to hours to days to months. The time course of the ellipticity change at 225 nm was used to follow the refolding/unfolding reactions at 4 °C. Slow unfolding kinetic traces (half times >200 s) were monitored upon the manual mixing of the folded protein with the buffer containing GdmCl to achieve the desired final protein concentration and the desired final denaturant concentration (typically 30 μM Lpp-56 monomer and 3−4 M GdmCl). For monitoring the folding kinetics, the protein was denatured by incubation for 4 min at 80 °C or in 6 M GdmCl at 4 °C. The folding reaction was initiated either by transferring the heat-denatured protein directly into the prethermostated cuvette mounted in the cell-holder of the spectropolarimeter or by rapid dilution into the prechilled buffer without GdmCl. The dead time of manual mixing was always <20 s, much shorter than the half time of the reaction under study. The temperature variation during manual mixing was estimated to be not larger than ±1 °C, and thermal equilibrium was established after about 3 min. The data were collected every 1 s in the initial phase and in larger intervals later in the course of reaction.

Stopped-Flow Kinetics. Fast unfolding experiments (half times <1 s) were performed with the π*-180 instrument (Applied Photophysics, Ltd.). Dead time was 1−2 ms; the optical path was 1 cm; the slits were set to 4 nm; and the detection wavelength was 225 nm. Unfolding was triggered by rapid mixing of protein solution in PBS with a buffer containing various concentrations of GdmCl at a mixing ratio of 1:10 and a flow rate of 5 mL s^{-1}. The final protein concentration was 5.4 μM (monomer equivalents), and the GdmCl concentrations varied between 4.53 and 5.45 M. At least three kinetic traces were averaged and used for further analysis.

Data Analysis. Equilibrium and kinetic data were analyzed according to the two-state-model assuming the presence of only the fully folded trimer (T) and the fully unfolded monomer (M) at each GdmCl concentration (*6*). The corresponding reaction scheme is as follows:

$$\mathrm{T} \underset{k_f}{\overset{k_u}{\rightleftharpoons}} 3\mathrm{M}$$

where k_u and k_f are the microscopic rate constants for unfolding and refolding, respectively. The unfolding equilibrium constant is defined in the usual way as $K_U = k_u/k_f$ and is related to the free energy change by $\Delta G = -RT \ln K_U$. For a given total concentration of the folded trimer, C_T, and defining f_U as the molar fraction of the unfolded protein (monomer), K_U is expressed by

$$K_U = \frac{27 {f_U}^3 {C_T}^2}{1 - f_U} \qquad (1)$$

f_U is calculated from the only real solution of eq 1

$$f_U = \sqrt[3]{A} - \frac{1}{3}\frac{k}{A} \qquad (2)$$

where

$$k = \frac{K_U}{27 C_T} \text{ and } A = k\left(\frac{1}{2} + \frac{\sqrt{3}\sqrt{4k + 27}}{18}\right)$$

The equilibrium unfolding data were analyzed by the linear extrapolation method (LEM) using an in-house written script for Mathematica (Wolfram Research, Inc.) for nonlinear least-squares optimization of the values for ΔG_U in the absence of the denaturant and $m_{eq} = d\Delta G_U/d[\mathrm{GdmCl}]$ as described elsewhere (*6*). The midpoint of denaturation, $[\mathrm{GdmCl}]_{1/2}$ was calculated from the maximum of the first derivative of the $df_U/d[\mathrm{GdmCl}]$ function.

Kinetic traces were analyzed by the numerical integration of the following equations:

$$\frac{d[M]}{dt} = -3k_f[M]^3 + 3k_u[T] \qquad (3a)$$

$$\frac{d[T]}{dt} = k_f[M]^3 - k_u[T] \qquad (3b)$$

Modeling according to a more complicated mechanism assuming the presence of a dimeric intermediate indicated a negligible fractional population of the dimer (less than 1%) and led to no improvement of the fitting statistics. If the simple two-state folding/unfolding mechanism is correct, the microscopic rates constants of folding and unfolding are expected to change exponentially with denaturant concentration according to the following equations:

$$\ln k_f = \ln k_f^{H_2O} + m_f[\mathrm{GdmCl}] \qquad (4a)$$

$$\ln k_u = \ln k_u^{H_2O} + m_u[\mathrm{GdmCl}] \qquad (4b)$$

The superscript H_2O refers to the rate constants in the absence of the denaturant, and m_f and m_u (in units of M^{-1}) describe the linear dependence of $\ln k_f$ and $\ln k_u$ on the denaturant concentration, respectively.

Numerical Simulations. Eqs 3a and 3b were used to simulate the time-dependent changes in the concentrations of the trimer and monomer in different GdmCl concentrations. The rate constants were calculated by interpolation of the functions given by eqs 4a and 4b and representing best fits to the experimental data (straight lines in Figure 4). The folding reactions were simulated starting with $[M] = 12\ \mu M$, and the unfolding reactions were simulated with $[T] = 4\ \mu M$ so that at infinite time the simulated fractional population of M and T at each GdmCl concentration should correspond to the fractional population of M and T observed in the equilibrium unfolding experiment, if the two-state character of folding/unfolding holds for Lpp-56. From the simulations, f_U was calculated as [M]/3[T].

RESULTS

The Lpp-56 protein is a stable trimer in solution (*3*). The mean residue ellipticity at 225 nm (MRE_{225}) in the micromolar concentration range is $\sim -34\,000$ deg cm^2 dmol^{-1}, indicating almost 100% helical content. The apparent midpoint of thermal denaturation is highly dependent on the scanning rate, and only a fraction of the helical signal is regained after fast cooling. However, if the same sample is incubated for prolonged times at 4 °C, MRE_{225} reaches ~90−95% of the initial value. These observations suggest, as has been reported earlier (*4*), that the thermal unfolding reaction is far from equilibrium during heating at experimentally accessible heating rates (down to 0.1 °C min^{-1}). Therefore, it is impossible to extract equilibrium thermodynamic parameters on Lpp-56 unfolding from thermal melting experiments.

Isothermal Solvent-Induced Denaturation. The thermodynamic stability of Lpp-56 was assessed at 4 °C by GdmCl-induced denaturation in PBS at pH 7. This temperature was selected for two reasons. First, Lpp-56 refolding exhibits anti-Arrhenius behavior, the refolding half times decreasing between 4 and 25 °C approximately by a factor of 10 (see Supporting Information). Second, the low-temperature minimizes the chance for bacterial contamination during the long incubation time. The conformational transitions were followed by monitoring MRE_{225}. The experiment was designed in the following way. The fully unfolded protein in 6 M GdmCl was diluted to a final 12 μM concentration (monomer equivalents) in PBS containing increasing amounts of the denaturant. Another series of samples was prepared by transferring the fully folded protein into PBS containing the desired final GdmCl concentration. After 18 h of incubation at 4 °C, the unfolding curves obtained from the two sets of samples were remarkably different (Figure 1). Although MRE_{225} at very low (below 0.4 M) and high (above 4 M) GdmCl were the same, the GdmCl concentrations where the protein was apparently half unfolded differed by almost 2 M. It follows that the equilibration between the folded trimer and the unfolded monomer is unusually slow not only at high temperatures but also at low temperatures in the presence of denaturant. The samples were incubated at 4 °C, and MRE_{225} was remeasured at time intervals until no further signal changes could be detected. This was the case around day 270 of incubation. An additional measurement was

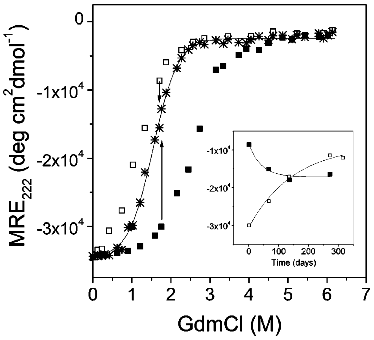


FIGURE 1: Equilibrium isothermal unfolding of Lpp-56 induced by GdmCl at 4 °C. Total protein concentration was 12 μM (monomer equivalents). The squares represent the MRE$_{225}$ measured after 18 h of incubation. Fully unfolded protein in 6 M GdmCl (open symbols) or folded protein in PBS (filled symbols) was diluted to the indicated final denaturant concentrations. MRE$_{225}$ was remeasured at time intervals, and the arrows indicate how the initial apparent unfolding curves converged after prolonged incubation. The asterisks represent MRE$_{225}$ after equilibrium has been reached (after ~270 days of incubation). The continuous line was computed by nonlinear least-squares fitting, assuming a two-state equilibrium between the folded trimer and the unfolded monomer. The data are best described by the following parameters: [GdmCl]$_{1/2}$, = 1.5 ± 0.1 M; $\Delta G_{\mathrm{U,eq}}^{\mathrm{H_2O}}$= 79 ± 10 kJ mol^{-1}; m_{eq} = 14.9 ± 2.5 kJ mol^{-1} M^{-1}. Inset: The time course of MRE$_{225}$ changes recorded in 1.7 M GdmCl.

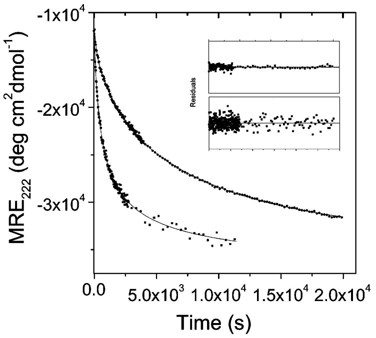


FIGURE 2: Refolding of Lpp-56. The time course of MRE$_{225}$ changes at 4 °C was followed after the rapid transfer of the completely unfolded protein to folding environment. Two representative kinetic traces collected at strong folding conditions are shown. Left trace: temperature jump from 80 °C to 4 °C. Final protein concentration was 24.7 μM (monomer equivalents). Right trace: hundred-fold dilution of Lpp-56 stock solution prepared in 6 M GdmCl and equilibrated at 4 °C. The final protein concentration was 11.2 μM (monomer equivalents), and the final GdmCl concentration was 0.06 M. The continuous lines are best fits according to the ternary reaction by numerical integration of eqs 3a and 3b. The residuals of the fits are shown in the inset.

performed at day 300 of incubation to verify the constancy of the signal. As seen in Figure 1, the ellipticity of the samples from the two sets of experiments at the same GdmCl concentration slowly converged over time. All data points were combined to calculate the stability of Lpp-56. The resulting unfolding curve could be reasonably described by a simple two-state unfolding of the folded trimer to the unfolded monomer. The free energy of unfolding at 4 °C, $\Delta G_{\mathrm{U,eq}}^{\mathrm{H_2O}}$, is 79 ± 10 kJ mol^{-1}. The midpoint of the unfolding transition, [GdmCl]$_{1/2}$, is 1.5 ± 0.1 M, and the dependence of ΔG_{unf} on the concentration of GdmCl, m_{eq}, is 14.9 ± 2.5 kJ mol^{-1} M^{-1} (Table 1).

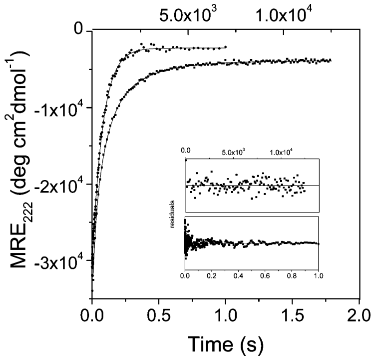


FIGURE 3: Unfolding of Lpp-56; time course of MRE$_{225}$ changes at 4° C after the rapid dilution of folded protein into a GdmCl-containing buffer. Upper trace and lower *x*-axis: unfolding in 5.45 M GdmCl measured by CD stopped-flow. The final protein concentration was 5.4 μM (monomer equivalents). Lower trace and upper *x*-axis: unfolding in 3.25 M GdmCl. The experiment was performed by manual mixing using a conventional CD spectropolarimeter. The final protein concentration was 31 μM (monomer equivalents). The continuous lines are best fits according to the ternary reaction by numerical integration of eqs 3a and 3b. The residuals of the fits are shown in the inset.

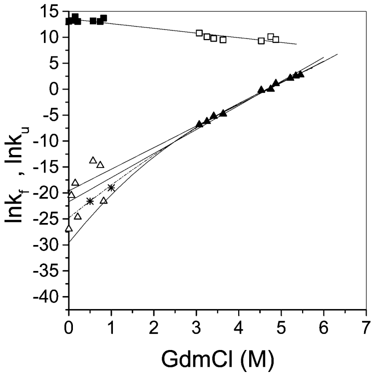


FIGURE 4: GdmCl dependence of the rate constants for refolding (■, □) and unfolding (▲, △). The rate constants k_f and k_u were calculated from the numerical integration of eqs 3a and 3b. The filled symbols represent data obtained at strong folding (<1 M GdmCl) and strong unfolding (>3 M GdmCl) conditions. Unfolding data were collected by manual mixing (3−4 M GdmCl) and by a CD stopped flow (>4.5 M GdmCl). The open symbols represent the natural logarithm of rate constants for unfolding and refolding extracted from analysis of kinetic traces, which were recorded at GdmCl concentrations where refolding and unfolding are dominating. Continuous lines are linear fits according to eqs 4a and 4b. All refolding data are best described with $k_f^{\mathrm{H_2O}}$= (7.6 ± 1.5) × 10^5 M^{-2} s^{-1} and m_f = 0.91 ± 0.08 M^{-1}. The upper line associated with unfolding data was calculated considering only data collected at strong unfolding conditions ($k_u^{\mathrm{H_2O}}$= (3.1 ± 1.0) × 10^{-9} s^{-1}; m_u = 4.16 ± 0.11 M^{-1}). The lower line describes all unfolding data ($k_u^{\mathrm{H_2O}}$= (3.8 ± 3.0) × 10^{-10} s^{-1}; m_u = 4.63 ± 0.50 M^{-1}). The asterisks are the natural logarithm of rate constants, which, according to numerical simulations, would explain the deviation between the equilibrium and kinetic data below ~1.2 M GdmCl (see Results and Figure 5 for details). The dotted line is a function smoothly connecting the simulated data with the directly measured data in the unfolding region. The intercept at 0 M GdmCl predicts $k_u^{\mathrm{H_2O}}$= 2 × 10^{-11} s^{-1}. The lowest continuous line illustrates a hypothetical dependence of ln k_u on [GdmCl]. Details are given in the text.

Folding/Unfolding Kinetics. The kinetic behavior of the Lpp-56 protein was studied at 4 °C by following the time

Table 1: Equilibrium and Kinetic Parameters Describing Folding of Lpp-56[a]

equilibrium parameters	kinetic parameters
$[\mathrm{GdmCl}]_{1/2}{}^{b}$	$k_f^{H_2O} \times 10^{-5}$
1.5 ± 0.1	6.3 ± 1.5^{c}
	6.6 ± 0.8^{d}
	$7.6 \pm 1.5^{e,f}$
	average
	6.8 ± 2.3
	$k_u^{H_2O} \times 10^{9}$
	$3.10 \pm 1.00^{f,g}$
	$0.38 \pm 0.30^{f,h}$
	0.02 ± 0.02^{i}
	$\sim 0.0001^{j}$
	m_f
	-0.91 ± 0.08^{f}
	m_u
	$4.16 \pm 0.11^{f,g}$
	$4.63 \pm 0.50^{f,h}$
$\Delta G_{U,eq}^{H_2O\,b}$	$\Delta G_{U,kin}^{H_2O\,k}$
79 ± 10	76 ± 15^{g}
	81 ± 20^{h}
	88 ± 20^{i}
	$\sim 100^{j}$
$m_{eq}{}^{b}$	$m_{kin}{}^{l}$
14.9 ± 2.5	11.7 ± 0.2^{g}
	12.7 ± 0.5^{h}

[a] In PBS at pH 7 and 4 °C. $[\mathrm{GdmCl}]_{1/2}$ in units of M; $k_f^{H_2O}$ in units of $M^{-2}\ s^{-1}$; $k_u^{H_2O}$ in units of s^{-1}; m_f and m_u in units of M^{-1}; $\Delta G_U^{H_2O}$ in units of kJ mol^{-1}; m_{eq} and m_{kin} in units of kJ $mol^{-1}\ M^{-1}$. [b] The parameters obtained by nonlinear optimization of a two-state model to the data (*6*). [c] Refolding in plain buffer; the mean of three experiments. [d] The mean of rate constants measured in 0−0.82 M GdmCl. [e] The linear fitting of data between 0 and 5.2 M GdmCl. [f] The parameters obtained by the linear fitting of eqs 4a or 4b. [g] The linear fitting of data between 3 and 5.5 M GdmCl. [h] The linear fitting of data between 0 and 5.5 M GdmCl. [i] With simulated curvature below 1 M GdmCl; dotted line in Figure 4. [j] Estimated upper limit (see text for details). [k] Calculated according to $\Delta G_U = -RT \ln(k_u/k_f)$. [l] Calculated according to $m_{kin} = RT(|m_f| + |m_u|)$.

course of the MRE_{225} change accompanying the formation/disruption of the helical structure. Figures 2 and 3 illustrate representative refolding and unfolding experiments. All kinetic traces were well reproduced when solving the differential equations describing the ternary monomer-to-trimer reaction (eqs 3a and 3b). Because the apparent refolding half times are long in the low-micromolar range of protein concentrations, refolding kinetic experiments could be performed at strong folding conditions (subdenaturing, 0−1 M GdmCl concentrations). Refolding was initiated either from the heat denatured state or form the chemically denatured state in 6 M GdmCl. The refolding rate constant measured in plain buffer is $k_f^{H_2O} = (6.3 \pm 1.5) \times 10^5\ M^{-2}\ s^{-1}$. Within error, k_f was independent of the GdmCl concentration below 1 M ($m_f = \mathrm{dln}\ k_f/\mathrm{d}[\mathrm{GdmCl}] = -0.004 \pm 0.45\ M^{-1}$). The average value of $k_f^{H_2O}$ from all 9 experiments is $(6.6 \pm 0.8) \times 10^5\ M^{-2}\ s^{-1}$. Still another estimate of $k_f^{H_2O}$ can be obtained if one assumes that the linear dependence of the refolding rate constant on the denaturant concentration holds in a broad range of [GdmCl] and considers k_f values obtained from the analysis of unfolding traces collected at strong denaturing conditions ([GdmCl] >3 M; Figure 4). Although, the fitting values of k_f in unfolding experiments are statistically underdetermined, the whole set of 16 data points can be reasonably described by a straight line ($R^2 = 0.913$). Extrapolation to 0 M GdmCl yields $k_f^{H_2O} = (7.6 \pm 1.5) \times 10^5\ M^{-2}\ s^{-1}$. Considering the stated errors, all three values agree well, and their average is $(6.8 \pm 2.3) \times 10^5\ M^{-2}\ s^{-1}$. From the slope of the straight line shown in Figure 4, one calculates $m_f = -0.91 \pm 0.08\ M^{-1}$.

The rates of unfolding were measured either in the stopped-flow instrument or by manual dilution of the folded protein into buffer containing increasing GdmCl concentrations (Figure 3). Both sets of data fit very well to each other and describe an exponential increase of the unfolding half times with increasing denaturant concentration in the tested range (Figure 4). The apparent rate constant of unfolding, $k_u^{H_2O}$, was estimated in two ways. The data collected at [GdmCl] >3 M are consistent with linear eq 4b ($R^2 = 0.994$), from which we calculate $k_u^{H_2O} = (3.1 \pm 1.0) \times 10^{-9}\ s^{-1}$ and $m_u = \mathrm{dln}\ k_u/\mathrm{d}[\mathrm{GdmCl}] = 4.16 \pm 0.11\ M^{-1}$. Because the extrapolation to 0 M GdmCl is a very long one and the assumption of eq 4b might not be generally valid, we assessed k_u values that were calculated as fitting parameters in the numerical analysis of refolding reactions at strong folding conditions according to eq 3a. As seen in Figure 4, and as expected, the confidence in these values is low. They do not exhibit any systematic dependence on [GdmCl] ($R^2 = 0.14$). The best linear fit of all data according to eq 4b results in $k_u^{H_2O} = (3.8 \pm 3.0) \times 10^{-10}\ s^{-1}$ and $m_u = \mathrm{dln}\ k_u/\mathrm{d}[\mathrm{GdmCl}] = 4.63 \pm 0.50\ M^{-1}$ ($R^2 = 0.935$).

The calculated rate constants for refolding and unfolding are listed in Table 1. From kinetic data, the free energy of unfolding at 4 °C, $\Delta G_{U,kin} = -RT \ln(k_u^{H_2O}/k_f^{H_2O})$, is 76 ± 15 and 81 ± 20 kJ mol^{-1}, $m_{kin} = RT(|m_f| + |m_u|)$ is 11.7 ± 0.2 and 12.7 ± 0.5 kJ $mol^{-1}\ M^{-1}$, depending on which estimate of $k_u^{H_2O}$ and m_u is used in the calculation. Although the errors are large, the correspondence between parameters derived from independent equilibrium and kinetic experiments is very good. Assuming simple refolding/unfolding kinetics and using the kinetic parameters from Table 1, one can calculate k_f and k_u and simulate the expected equilibrium concentrations of the folded trimer and the unfolded monomer at any GdmCl concentration. Figure 5 illustrates the overall similarity between the unfolding curve measured at equilibrium and the one derived from kinetic experiments. However, a closer inspection reveals that the extent of unfolding that is predicted from the linear extrapolation of k_u is larger than that measured in equilibrium below approximately 1 M GdmCl. Numerical simulations demonstrate that the correspondence between equilibrium and kinetic data is significantly improved if one assumes a downward concave curvature of the ln k_u versus [GdmCl] plot (asterisks and dotted line in Figure 4).

DISCUSSION

We have measured the stability and the kinetic behavior of the trimeric coiled-coil core of the *E. coli* outer membrane lipoprotein (Lpp-56). Because both refolding and unfolding are slow in comparison to those in many other proteins studied to date, estimating the thermodynamic stability and the rates of formation and disruption of the native structure is not a trivial task. Because of the almost complete kinetic

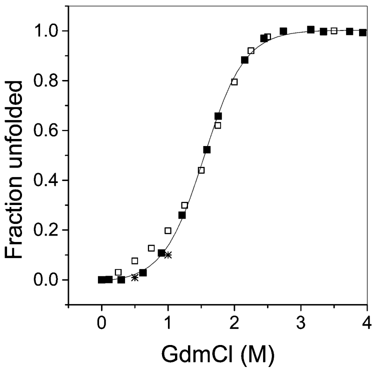


FIGURE 5: GdmCl dependence of the equilibrium distribution between the folded trimer and the unfolded monomer. The filled symbols represent the fraction of unfolded Lpp-56 calculated from the equilibrium unfolding experiment (see Figure 1). Open symbols are the results of numerical simulations according to eqs 3a and 3b as described in Experimental Procedures. The asterisks are f_U values calculated from simulations in which k_f were taken from the best linear fit of eq 4a to the folding data in Figure 4, whereas k_u values were set to 5.6×10^{-9} s^{-1} (0.5 M GdmCl) and 4×10^{-10} s^{-1} (1 M Gdm), instead of 4×10^{-9} s^{-1} and 3.9×10^{-9} s^{-1}, respectively, which correspond to the best linear fit of eq 4b to the unfolding data in Figure 4.

irreversibility, the application of the van't Hoff formalism to extract thermodynamic information from spectroscopic and calorimetric thermal unfolding data is impossible (*4*). We assessed the stability of Lpp-56 at low temperature by using GdmCl as the denaturing agent. ΔG_U of unfolding was calculated by the standard linear extrapolation method (LEM) in the limit of equilibrium two-state folding/unfolding transition. In view of the known kinetic peculiarity of the protein, the approach requires justification. In a reversible process, the same partitioning of molecular species is expected at equilibrium, regardless of the initial state of the system. This was indeed the case with Lpp-56. At each GdmCl concentration, MRE_{225} converged to the same value (within the error of concentration determination and instrumental noise) if equilibration was attained starting either from the fully denatured state or from the fully folded state. As to the validity of the two-state approximation, the presence of intermediate states cannot be completely ruled out on the sole basis of denaturant-induced unfolding data. For example, the observed ellipticity in the transition region could contain contribution from intermediates with partially developed helical structure. However, the unfolding curve shown in Figure 1 is symmetrical around the midpoint of unfolding and exhibits no detectable kinks. It is unlikely that the population of major intermediates would shift smoothly with the variation in GdmCl concentration. Alternatively, intermediates without helical structure might escape detection by CD spectroscopy. Because the Lpp-56 monomers are unstructured in isolation and the folded trimer is stabilized by nonpolar contacts at the interface between the helices, the presence of nonhelical, sufficiently stable states appears very improbable. Light scattering experiments in aqueous solutions at higher temperatures failed to detect dimeric intermediates (*4*). A strong argument in favor of the validity of the two-state approximation comes from kinetic data. The refolding and unfolding traces can be modeled with simple trimolecular refolding and unimolecular unfolding reactions, respectively. The limbs of the chevron plot shown in Figure 4 are reasonably linear. There is a very good agreement between ΔG_U and m values calculated from equlibrium and kinetic data (Table 1) as well as between unfolding curves derived from equilibrium and kinetic data (Figure 5).

There are two published studies dealing with the thermodynamic stability of Lpp-56. Shu et al. estimated the unfolding free energy as 86 kJ mol^{-1} from GdmCl denaturation (*3*). However, this number was derived at nonequilibrium conditions and, therefore, does not reflect the true free energy change of unfolding. The numerical similarity between ΔG_U obtained in ref 3 (*3*) and that in our study is perhaps a fortuitous event. More recently, DP suggested that the genuine thermodynamic stability of Lpp-56 at 4 °C is much higher, on the order of 140 kJ mol^{-1} (*4*). The difference between the latter number and our own estimate of ∼80 kJ mol^{-1} is striking. Both analyses are based on the two-state model of unfolding. It must be noted, however, that the two studies differ in one fundamental aspect. DP used heat, whereas we used GdmCl to induce a shift of the conformational equilibrium between the native and the denatured state and modulate the rates of folding and unfolding. Heat and chemical denaturants have been widely used to examine the energetics of proteins, but the physical origin of their action as denaturing agents is quite different. Discrepancies between ΔG_U derived from thermal and chemical denaturation are documented (*7*−*10*). The applicability and pitfalls of LEM as a tool to quantify protein stability have been discussed from theoretical and experimental points of view (*11*−*14*).

The height of the activation barrier for folding, $\Delta G_{\ddagger-D}$, is rigorously defined because the refolding rate was measured directly in benign buffer at low temperature, and we, along with DP, obtained very similar values for k_f. Moreover, the folding rate variation upon increasing the GdmCl concentration and increasing the temperature is relatively weak (Figure 4 in this article, and Figure 5 in ref 4 (*4*)). Given that the two-state model of Lpp-56 unfolding holds reasonably well for both thermal- and denaturant-induced unfolding, the disparity in ΔG_U values has to be attributed to different heights of the unfolding activation barrier, $\Delta G_{\ddagger-N}$, which was estimated using long extrapolation. Indeed, k_u was directly measured only at strong unfolding conditions between 3 and 5.5 M GdmCl and between 60 and 70 °C. In principle, the linearity of ln k_u versus [GdmCl] or ln k_u versus $1/T$ plots would indicate that $\Delta G_{\ddagger-N}$, that is, the energetic difference between the native and the transition state, changes smoothly over the considered range of denaturant concentrations or temperatures. However, linearity outside the experimentally accessible range of solvent conditions is not granted a priori. As demonstrated in Figures 4 and 5, the data indicate a possible downward concave curvature of the ln k_u versus [GdmCl] plot (asterisks and dotted line in Figure 4).

Among alternatives, the simplest and perhaps the most intuitive explanation for the predicted deviation from linearity comes from considering electrostatic effects. GdmCl is a salt and attenuates electrostatic interactions by screening the Coulombic attraction between opposite charges and/or by altering the desolvation contributions to the total electrostatic free energy. The effect is concentration-dependent, implying that a pronounced change of the electrostatic component of ΔG_U is expected between 0 and 1 M GdmCl, at which salt concentration, charge−charge interactions are usually assumed to be insignificant. If the native state is stabilized by electrostatics, the addition of salt will increase its free energy and, therefore, $\Delta G_{\ddagger-\mathrm{N}}$ will abruptly decrease between 0 and 1 M, yet linear extrapolation from much higher GdmCl concentrations will fail to detect the effect. Alternatively, GdmCl might stabilize the transition state by abolishing unfavorable electrostatic interactions (or in some other way). In view of the very weak variation of k_f between 0 and 1 M GdmCl (Figure 4), we render this possibility unlikely.

Indeed, electrostatic effects appear to be an important factor governing the the conformational stability of Lpp-56. There are close ionic contacts seen in the crystal structure, some of them forming salt bridge networks (*3*). DP convincingly demonstrated that the stability of Lpp-56 is much lower at pH 3, where at least most of the salt bridges are disrupted. In fact, ΔG_U at pH 3 is almost half of that at neutral pH according to their estimate and is very close to what we have found. We are currently investigating the electrostatic contribution to Lpp-56 stability by computational methods. In the following, we argue, however, that the lower ΔG measured by GdmCl-induced unfolding cannot be entirely explained by the electrostatic effect of GdmCl.[2]

Charge screening (or other attenuation of electrostatics) by GdmCl is expected to induce a steep decrease of $|m_\mathrm{kin}|$ (and $|m_\mathrm{eq}|$) between 0 and 1 M denaturant. Because $m_\mathrm{kin} = RT(|m_\mathrm{f}| + |m_\mathrm{u}|)$ and the dependence of $\ln k_\mathrm{f}$ on GdmCl, m_f, is quite small (see Figure 4), it is m_u that largely accounts for the experimentally observed m_kin. It follows that a pronounced downward curvature of the ln k_u versus [GdmCl] plot will result. An arbitrarily drawn ln k_u dependence is shown in Figure 4 (lowest continuous line). It smoothly deviates from linearity below ~1.5 M GdmCl. The actual curvature of this function is indeed not known, yet there are two important constraints. First, the deviation from linearity is very unlikely to be pronounced above 1−1.5 M GdmCl (*8*, *9*, *15*, *16*). Second, the slope around 0 M GdmCl cannot exceed some limiting value because $|m_\mathrm{u}| \leq m_\mathrm{eq}/RT - |m_\mathrm{f}|$. Setting $m_\mathrm{f} = 0$, this limit is given by some physically reasonable m_eq value. For a protein of the size of Lpp-56, the expected m value is on the order of 20−25 kJ mol^{-1} M^{-1}, if unfolding is complete (*13*), and therefore, $|m_\mathrm{u}|$ cannot exceed 9−11 M^{-1}. The line shown in Figure 4 was drawn with $|m_\mathrm{u}| = 10\ \mathrm{M}^{-1}$ between 0 and 1 M GdmCl. The exercise leads to the conclusion that the true unfolding rate constant in the absence of GdmCl is on the order of 10^{-13} s^{-1} as a limit, implying that genuine Lpp-56 stability is on the order of 100 kJ^{-1} mol^{-1}. Hence, the linear extrapolation procedure underestimates the Lpp-56 stability by 10−20 kJ mol^{-1}. The differences in protein stability (including the stability of coiled coils) calculated from GdmCl or urea/thermal unfolding have been invariably estimated in the range 5−15 kJ mol^{-1} (*7*−*10*, *15*, *17*−*20*).

The above discussion does not intend to invalidate the experimental results of DP in any way. It only illustrates that arguments based solely on salt effects seems not to reconcile the ΔG_U that we found from chemical unfolding with that found by DP from thermal unfolding. The two studies provide contradictory views about the nature of the transition state as well. DP found the main change in ΔC_p to occur between the unfolded monomeric state and the transition state, whereas the heat capacity difference between the native trimer and the transition state is negligible. As far as the ΔC_p of the protein, conformational changes are mostly caused by changes in the hydration of apolar groups, this observation suggests high structural similarity between the native and transition states. In contrast, we found that Lpp-56 traverses an unstructured and highly solvent-accessible high energy state. This follows from the very low Tanford β_T value, which reflects the buried surface area in the transition state relative to the native state. ($\beta_\mathrm{T} = 1 - |m_\mathrm{f}|/(|m_\mathrm{f} + m_\mathrm{u}|)$ ~0.15; see Table 1). The transition state in the folding of short dimeric coiled coils has been found to be either significantly structured or largely unstructured, depending on the experimental approach (*21*−*24*). It has been proposed that the high-temperature transition state moves toward the native state in accord with the Hammond postulate (*25*). It is possible that the structural properties of, and thereby the energetic differences between, the ground states and the activated state vary in different ways with the temperature and denaturant concentration, and thus, different unfolding pathways have been sampled in the two studies.[3]

The role of electrostatic interactions in general and of salt bridges in particular to coiled coil stability has been a subject of controversy for some time. Both stabilization and destabilization by surface-exposed ion pairs have been documented (*18*, *20*, *26*). To our knowledge, there are no cases where salt bridges dominate the structural energetics of coiled coils, which are mainly stabilized by the packing of nonpolar side chains at the interhelical interface. It appears that charge−charge attractions and repulsions are factors that modulate the rate of association-coupled refolding (*27*), assist (and maintain) the in-register alignment and the directionality of the super-helix (*28*), and govern the specificity of oligomerization (*29*). In this respect, our finding that the rate of Lpp-56 refolding is much less affected by the salt nature of GdmCl than the rate of unfolding is surprising. It appears that ionic contacts genuinely stabilize the folded conformation of this trimeric coiled-coil. It must be noted, however, that the term salt bridge is usually employed when referring to the close proximity of two oppositely charged groups. The nonpolar moiety of ionizable side chains might participate in hydrophobic packing as well. This is the case in Lpp-56, where residues occupying the canonical hydrophobic *a* and *d* heptad positions pack not only against each other but also against the charged residues occupying the *e* and *g* positions to complete the hydrophobic core (*3*). Therefore, the disruption of a salt bridge by pH or salt would also alter packing and would lead to an indirect destabilization. Indeed, at pH 3, DP found a significant increase in the temperature

---

[2] The similarity between ΔG_U measured at pH 3 and ΔG_U measured in GdmCl cannot be taken at face value as representing the same electrostatic effect. The lowering of pH will induce the protonation of acidic groups and thereby will cause the disruption of salt bridges. However, the degree of protonation depends on the actual (and a priori unknown) $\mathrm{p}K_\mathrm{a}$ values of individual Asp and Glu side chains. Furthermore, the denatured state will be highly positively charged. The situation is different in GdmCl. If, for simplicity, we assume the primary electrostatic effect of GdmCl to be the screening of charge−charge interactions, these interactions will be screened not only in the native state but also in the denatured state.

[3] It is worth noting that ΔC_p (11 J K^{-1} (mol res^{-1})) and m_eq (62 J $(\mathrm{mol\ res})^{-1}$ M^{-1}) for Lpp-56 unfolding are both much lower than the mean values for globular proteins (58 J K^{-1}(mol res^{-1}) and 98 J(mol $\mathrm{res})^{-1}$ M^{-1}, respectively).

dependence of the partial molar heat capacity of native Lpp-56 and a much lower enthalpy and entropy, characterizing the transition from the folded to the activated state.

Why is Lpp-56-folding and refolding so slow? The folding of an oligomeric coiled coil is tightly coupled to the association of polypeptides, which are intrinsically devoid of stable tertiary structure in isolation. For a trimeric coiled coil, the simultaneous association of three chains is very unlikely from a statistical point of view. Furthermore, folding could be slow if the folding mechanism involves the formation of a nucleation site by association of monomeric chains that are partly helical and/or in-register, that is, in the correct orientation allowing structure propagation. The presence of a sequence triggering the folding of dimeric coiled coils has been suggested (*30*−*32*). Interestingly, the AGADIR algorithm predicts for the isolated Lpp-56 chains a 15−25% helical propensity localized in the *C*-terminal stretch encompassing the alanine zipper (Supporting Information). Along the same lines, structure consolidation toward the optimal, in-register alignment of the monomers might involve the sliding of the polypeptide chains within an initial associated state. Of note is the fact that all *a* and *d* positions participating in the Lpp-56 hydrophobic core are occupied by nonpolar residues, whereas most naturally occurring coiled coils contain buried polar residues (*33*). Such polar clusters have been demonstrated to warrant the proper orientation of coiled coil monomers and to govern the dimerization specificity. It is likely that they may also confer a mechanism speeding up folding by providing buoys in the hydrophobic core. Although speculative at present, this suggestion is corroborated by our studies at pH 5. We tested the effect of the replacement of alanine 37 (*d* heptad position) and alanine 41 (*a* heptad position) by polar glutamine. The two variants were more stable in thermal unfolding experiments (apparent T_m shifts of 8 and 10 °C, respectively) and refolded faster than wild-type Lpp-56 by a factor of 10 (Supporting Information).

The very slow unfolding rate is most likely a combination of different factors. Unfolding accompanied by the simultaneous dissociation of the three chains will require the simultaneous disruption of the packing interactions in the extended hydrophobic core, the breakage of (apparently strong) salt bridges, and the opening of the *N*- and *C*-terminal caps. Such a concerted event is expected to occur rather infrequently. The unfolding half times at benign conditions are 10^2-10^5 years (this study) and 10^{12} years (derived from the data of DP). Both numbers are difficult to comprehend. Very slow unfolding is not exceptional, and plausible arguments have been put forward in favor of the existence of evolutionary pressure for increasing the kinetic stability of certain classes of proteins (*34*, *35*). To our knowledge, however, Lpp-56 is the first coiled coil for which a high unfolding kinetic barrier has been experimentally demonstrated.

ACKNOWLEDGMENT

We thank Dr. Min Lu for providing the protein used in this study, Dr. Daniel Nettels for help with Mathematica scripts, and Dr. Christine Berger Sprecher for valuable discussions.

SUPPORTING INFORMATION AVAILABLE

Figures describing thermal unfolding and refolding kinetics of wild-type Lpp-56 and two Ala-to Gln variants at pH 5. This material is available free of charge via the Internet at http://pubs.acs.org

REFERENCES

1. Goodsell, D. S., and Olson, A. J. (1993) Soluble-proteins − size, shape and function, *Trends Biochem. Sci. 18*, 65−68.
2. Lupas, A. N., and Gruber, M. (2005) The structure of alpha-helical coiled coils, *Adv. Protein. Chem. 70*, 37−78.
3. Shu, W., Liu, J., Ji, H., and Lu, M. (2000) Core structure of the outer membrane lipoprotein from *Escherichia coli* at 1.9 A resolution, *J. Mol. Biol. 299*, 1101−1112.
4. Dragan, A. I., Potekhin, S. A., Sivolob, A., Lu, M., and Privalov, P. L. (2004) Kinetics and thermodynamics of the unfolding and refolding of the three-stranded alpha-helical coiled coil, Lpp-56, *Biochemistry 43*, 14891−14900.
5. Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins, *Biochemistry 6*, 1948−1954.
6. Marti, D. N., Bjelic, S., Lu, M., Bosshard, H. R., and Jelesarov, I. (2004) Fast folding of the HIV-1 and SIV gp41 six-helix bundles, *J. Mol. Biol. 336*, 1−8.
7. Ferreon, A. C. M., and Bolen, D. W. (2004) Thermodynamics of denaturant-induced unfolding of a protein that exhibits variable two-state denaturation. *Biochemistry 43*, 13357−13369.
8. Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (1996) A model-independent, nonlinear, extrapolationprocedure for the characterization of protein folding energetics from solvent-denaturation data, *Biochemistry 35*, 14689−14702.
9. Makhatadze, G. I. (1999) Thermodynamics of protein interactions with urea and guanidinium hydrochloride, *J. Phys. Chem. B 103*, 4781−4785.
10. Pfeil, W. (2001) *Protein Stability and Folding: A Collection of Thermodynamic Data.*, Springer-Verlag, New York.
11. Alonso, D. O. V., and Dill, K. A. (1991) Solvent denaturation and stabilization of globular proteins, *Biochemistry 30*, 5974−5985.
12. Johnson, C. M., and Fersht, A. R. (1995) Protein stability as a function of denaturant concentration: The thermal-stability of barnase in the presence of urea, *Biochemistry 34*, 6795−6804.
13. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m-values and heat-capacity changes: Relation to changes in accessible surface-areas of protein unfolding, *Protein Sci. 4*, 2138−2148.
14. Yao, M., and Bolen, D. W. (1995) How valid are denaturant-induced unfolding free-energy measurements: Level of conformance to common assumptions over an extended range of ribonuclease A stability, *Biochemistry 34*, 3771−3781.
15. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I., and Sanchez-Ruiz, J. M. (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge−charge interactions to protein stability, *Biochemistry 38*, 8138−8149.
16. Plaxco, K. W., and de los Rios, M. A. (2005) Apparent Debye−Huckel electrostatic effects in the folding of a simple, single domain protein. *Biochemistry 44*, 1243−1250.
17. Bolen, D. W., and Yang, M. (2000) Effects of guanidine hydrochloride on the proton inventory of proteins: Implications on interpretations of protein stability, *Biochemistry 39*, 15208−15216.
18. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Protein denaturation with guanidine-hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions, *Protein Sci. 3*, 1984−1991.
19. Perez-Jimenez, R., Godoy-Ruiz, R., Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (2004) The efficiency of different salts to screen charge interactions in proteins: A Hofmeister effect? *Biophys. J. 86*, 2414−2429.
20. Phelan, P., Gorfe, A. A., Jelesarov, I., Marti, D. N., Warwicker, J., and Bosshard, H. R. (2002) Salt bridges destabilize a leucine zipper designed for maximized ion pairing between helices, *Biochemistry 41*, 2998−3008.
21. Bhattacharyya, R. P., and Sosnick, T. R. (1999) Viscosity dependence of the folding kinetics of a dimeric and monomeric coiled coil, *Biochemistry 38*, 2601−2609.

22. Bosshard, H. R., Durr, E., Hitz, T., and Jelesarov, I. (2001) Energetics of coiled coil folding: The nature of the transition states, *Biochemistry 40*, 3544−3552.
23. Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A., and Sosnick, T. R. (1999) Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking, *Proc. Natl. Acad. Sci. U.S.A. 96*, 10699−10704.
24. Sosnick, T. R., Jackson, S., Wilk, R. R., Englander, S. W., and DeGrado, W. F. (1996) The role of helix formation in the folding of a fully alpha-helical coiled coil, *Proteins: Struct., Funct., Genet. 24*, 427−432.
25. Otzen, D. E., and Oliveberg, M. (2004) Correspondence between anomalous m- and delta Cp-values in protein folding, *Protein Sci. 13*, 3253−3263.
26. Yu, Y., Monera, O. D., Hodges, R. S., and Privalov, P. L. (1996) Ion pairs significantly stabilize coiled-coils in the absence of electrolyte, *J. Mol. Biol. 255*, 367−372.
27. Wendt, H., Leder, L., Harma, H., Jelesarov, I., Baici, A., and Bosshard, H. R. (1997) Very rapid, ionic strength-dependent association and folding of a heterodimeric leucine zipper, *Biochemistry 36*, 204−213.
28. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in 2-stranded alpha-helical coiled-coils, *Biochemistry 33*, 3862−3871.
29. Yu, Y. B. (2002) Coiled-coils: Stability, specificity, and drug delivery potential. *Adv. Drug Delivery Rev. 54*, 1113−1129.
30. Kammerer, R. A., Jaravine, V. A., Frank, S., Schulthess, T., Landwehr, R., Lustig, A., Garcia-Echeverria, C., Alexandrescu, A. T., Engel, J., and Steinmetz, M. O. (2001) An intrahelical salt bridge within the trigger site stabilizes the GCN4 leucine zipper, *J. Biol. Chem. 276*, 13685−13688.
31. Steinmetz, M. O., Stock, A., Schulthess, T., Landwehr, R., Lustig, A., Faix, J., Gerisch, G., Aebi, U., and Kammerer, R. A. (1998) A distinct 14 residue site triggers coiled-coil formation in cortexillin I, *EMBO J. 17*, 1883−1891.
32. Zitzewitz, J. A., Ibarra-Molero, B., Fishel, D. R., Terry, K. L., and Matthews, C. R. (2000) Preformed secondary structure drives the association reaction of GCN4-p1, a model coiled-coil system, *J. Mol. Biol. 296*, 1105−1116.
33. Akey, D. L., Malashkevich, V. N., and Kim, P. S. (2001) Buried polar residues in coiled-coil interfaces, *Biochemistry 40*, 6352−6360.
34. del Pino, I. M. P., Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (2000) Lower kinetic limit to protein thermal stability: A proposal regarding protein stability in vivo and its relation with misfolding diseases, *Proteins: Struct., Funct., Genet. 40*, 58−70.
35. Jaswal, S. S., Truhlar, S. M. E., Dill, K. A., and Agard, D. A. (2005) Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases, *J. Mol. Biol. 347*, 355−366.

BI0608156

Supporting Information for World Wide Web Edition

Stability and Folding/Unfolding Kinetics of The Homotrimeric Coiled Coil Lpp-56[†]

Saša Bjelić, Andrey Karshikoff[‡]. and Ilian Jelesarov[*]

NH_2-SSNAKIDQLSSDVQTLNAKVDQLSNDVNAMRSDVQAAKDDAARANQRLDNMATKYR-COOH

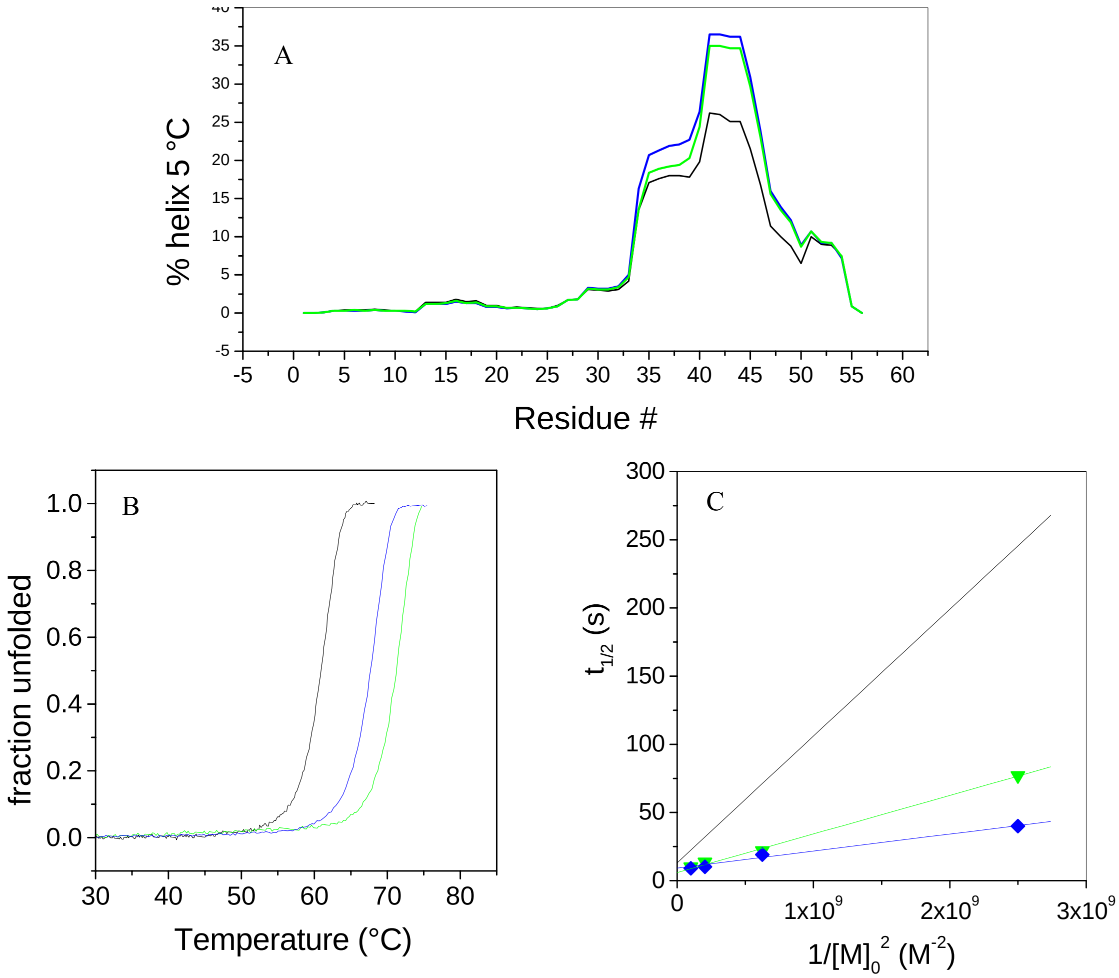


Figure S2. Effect of Ala-to-Gln mutations to the stability and folding kinetics of Lpp-56 at 4 °C and pH 5. On top, the sequence of Lpp-56 is given. Alanine 37 (*d* heptad position) and alanine 41 (*a* heptad position) were mutated to polar glutamine (underlined). **A,** Predicted helical content of wild type Lpp-56 (black), A37Q (blue) and A41Q (green). The calculation was performed with the standard setting parameters of AGADIR (http://www.embl-heidelberg.de/Services/serrano/agadir). **B,** The apparent fraction of unfolded peptide is plotted as a function of temperature for 100 μM concentration of wild type Lpp-56 (black), A37Q (blue) and A41Q (green). Scanning rate was 1 °C/min. **C**, Concentration dependence of the refolding half-times of wild type Lpp-56 (black), A37Q (blue) and A41Q (green). Experiments were performed by temperature jump from 70 °C or 80 °C to 4 °C in cocktail buffer, pH 5. Protein concentrations were 20, 40, 70 and 100 μM (monomer equivalents). The linearity of the plots indicates that refolding under strong folding conditions obeys the kinetics of a tri-molecular reaction. From the slopes, the following folding rate constants were calculated: 2.7×10^6 M^{-2} s^{-1} (WT), 8.6×10^6 M^{-2} s^{-1} (A41Q) and 2.4×10^7 M^{-2} s^{-1} (A37Q).

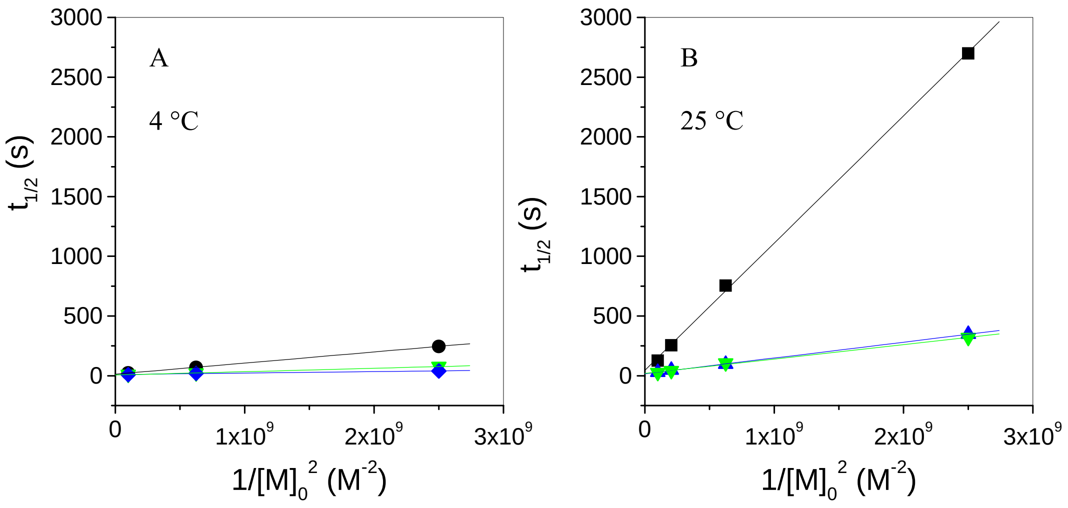


Figure S3. Concentration dependence of the refolding half-times of wild type Lpp-56 (black), A37Q (blue) and A41Q (green) as a function of the temperature. A, Data et 4 °C. B, Data at 25 °C. Experiments were performed by temperature jump from 70 °C or 80 °C to 4 °C in cocktail buffer, pH 5. Protein concentrations were 20, 40, 70 and 100 μM (monomer equivalents).

2.3 Electrostatic contribution to the thermodynamic and kinetic stability of the homotrimeric coiled coil Lpp-56: A computational study

Saša Bjelić and Silke Wieninger and Ilian Jelesarov and Andrey Karshikoff

Article published in Proteins 70: 810-822

WILEY InterScience®
DISCOVER SOMETHING GREAT

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Electrostatic contribution to the thermodynamic and kinetic stability of the homotrimeric coiled coil Lpp-56: A computational study

Saša Bjelić,[1] Silke Wieninger,[2,3] Ilian Jelesarov,[1] and Andrey Karshikoff[2]*

[1] Biochemisches Institut der Universität Zürich, Winterthurerstr, Zürich, Switzerland
[2] Department of Biosciences and Nutrition, Karolinska Institute, Stockholm, Sweden
[3] ERASMUS programme student on leave from University of Bayreuth

ABSTRACT

The protein moiety of the Braun's* E. coli *outer membrane lipoprotein (Lpp-56) is an attractive object of biophysical investigation in several aspects. It is a homotrimeric, parallel coiled coil, a class of coiled coils whose stability and folding have been studied only occasionally. Lpp-56 possesses unique structural properties and exhibits extremely low rates of folding and unfolding. It is natural to ask how the specificity of the structure determines the extraordinary physical chemical properties of this protein. Recently, a seemingly controversial data on the stability and unfolding rate of Lpp-56 have been published (Dragan* et al.*, Biochemistry 2004;43: 14891–14900; Bjelic* et al.*, Biochemistry 2006;45:8931–8939). The unfolding rate constant measured using GdmCl as the denaturing agent, though extremely low, was substantially higher than that obtained on the basis of thermal unfolding. If this large difference arises from the effect of screening of electrostatic interactions induced by GdmCl, electrostatic interactions would appear to be an important factor determining the unusual properties of Lpp-56. We present here a computational analysis of the electrostatic properties of Lpp-56 combining molecular dynamics simulations and continuum* pK *calculations. The pH-dependence of the unfolding free energy is predicted in good agreement with the experimental data: the change in* Δ*G between pH 3 and pH 7 is* $\sim$*60 kJ mol*$^{-1}$*. The results suggest that the difference in the stability of the protein observed using different experimental methods is mainly because of the effect of the reduction of electrostatic interactions when the salt (GdmCl) concentration increases. We also find that the occupancy of the interhelical salt bridges is unusually high. We hypothesize that electrostatic interactions, and the interhelical salt bridges in particular, are an important factor determining the low unfolding rate of Lpp-56.

Proteins 2008; 70:810–822.

***Key words:** electrostatic interactions; protein stability; unfolding kinetics; molecular dynamics; salt bridge; coiled coil.*

INTRODUCTION

The coiled coil is a ubiquitously encountered structural motif in proteins.[1] Amino acid stretches bearing direct repetition of the *abcdef* heptad pattern form two-, three-, or higher-order superhelices. The strandedness (degree of oligomerisation), the orientation (parallel vs. antiparallel), and the registry (in-register vs. out-of-register) are dictated by the nature of mostly aliphatic side chains occupying positions *a* and *d*, in an intimate interplay with the nature of mostly charged side chains occupying positions *e* and *g*. Given the apparent simplicity of the motif, coiled coils have attracted attention as a model for investigating the sequence–structure–energy relationships in protein folding. However, the main body of knowledge about coiled coil folding stems from investigations of short, dimeric, three- to five-heptad long species. Higher-order coiled coils are less well characterized.

The 56 amino acid long protein moiety of the *E. coli* outer membrane protein (henceforth referred to as Lpp-56) is a parallel, in-register, trimeric coiled coil.[2] The physiological importance of this protein in maintaining the structural integrity of the *E. coli* cell wall has been reviewed.[3] Recently, we have determined the equilibrium stability and the rates of unfolding and refolding of Lpp-56 at pH 7.[4] In this study we used GdmCl to shift the equilibrium between folded trimer and unfolded monomer, and to modulate the refolding and unfolding rates. The unfolding free energy, ΔG^{Gdm}, obtained by extrapolation to zero denaturant according to the linear extrapolation method is 79 ± 10 kJ mol^{-1}. A substantially larger value, $\Delta G^{\mathrm{therm}} = 137$ kJ mol^{-1}, has been determined

---

Grant sponsor: Swiss National Science Foundation; Grant number: 3100A0-100197.
*Correspondence to: Andrey Karshikoff, Ph.D., Dr.Sci., Karolinska Institutet, Dep. Biosciences and Nutrition, S-14157 Huddinge, Sweden. E-mail: aka@csb.ki.se
Received 17 January 2007; Revised 20 March 2007; Accepted 13 April 2007
DOI: 10.1002/prot.21585

from calorimetric experiments.[5] Interestingly, $\Delta G^{\mathrm{therm}}$ at pH 3, where salt bridges are believed to be disrupted by protonation of acidic side chains, is on the order of 70 kJ mol^{-1}.[5] Since GdmCl is a salt it is intuitive to assume that the discrepancy can be attributed to screening of electrostatic interactions when the unfolding free energy is evaluated by linear extrapolation of data collected at high salt (GdmCl) conditions, leading to a severe underestimation of ΔG^{Gdm}. However, we reasoned[4] that if Lpp-56 obeys the two-state unfolding model (neither we, nor Dragan *et al.* have detected intermediate states) the contribution of electrostatic interactions that are eliminated by salt should be between 10 and 20 kJ mol^{-1}. The latter value is pretty close to previous estimates[6–11] of the magnitude of the electrostatic contribution to protein stability, but is rather low to explain the difference in the unfolding free energy obtained by us and by Dragan *et al.*

One striking property of Lpp-56 is its extremely low unfolding rate. Our own estimate[4] of the unfolding rate constant is on the order of 10^{-10} to 10^{-13} s^{-1}. The data of Dragan *et al.*[5] predict an even lower unfolding rate constant, on the order of 10^{-21} s^{-1}. Both estimates demonstrate an extremely high kinetic stability of Lpp-56, yet the difference appears too large to be explained by the charge screening effect of GdmCl used in our experiments. Nevertheless, based on the steep dependence of the activation unfolding energy on GdmCl concentration at pH 7 and the much faster unfolding at pH 3, we speculated that there is a large electrostatic contribution to the free energy barrier for unfolding.

Indeed, each Lpp-56 polypeptide contains eight acidic side chains (exclusively aspartic acid) and eight basic side chains (four lysines and four arginines). In addition, the C-terminal carboxylates and the three tyrosines (one per chain) are potentially capable to participate in salt bridge formation. As in other coiled coils, reflecting the repetitive heptad organization of the molecule, the interhelical salt bridges form a system of rings girdling the three-helix bundle of Lpp-56 along its length (see Fig. 1).

These peculiarities of Lpp-56, together with the large number of potential salt bridges seen in the crystal structure[2] motivated us to investigate the role of electrostatic interactions in the stabilization of the native structure of Lpp-56. To our knowledge, there are no published experimental data on ionization constants of the titratable groups or a potentiometric titration curve of Lpp-56, which can facilitate the analysis of electrostatic interactions. Therefore, our investigation essentially relies on theoretical prediction of ionization equilibria. This creates some difficulties in the quantitative assessment of the results, especially results concerning the thermodynamic stability of the protein. In spite of this, we extract information about the magnitude of the screening effect of salt on electrostatic interactions. Furthermore, we demonstrate the presence of unusually stable salt bridges and formulate a hypothesis explaining the extremely low unfolding rate of Lpp-56.

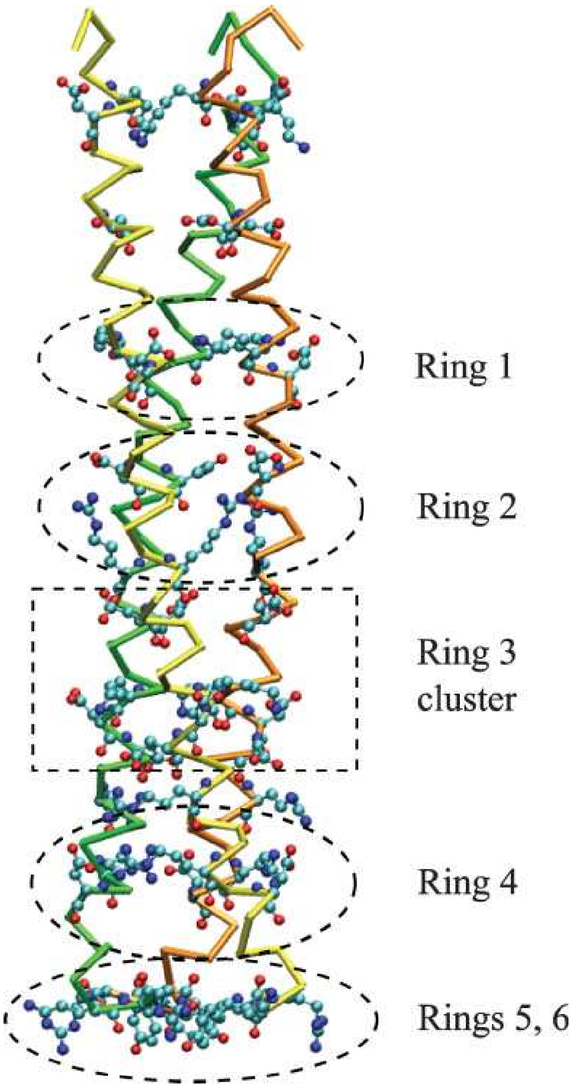


Figure 1

Titratable side chains in Lpp-56. The groups forming salt bridges are grouped in rings along the rod-like molecule (see also Table III). The description of the rings is given in some detail in section "Role of the salt bridges in unfolding kinetic of Lpp-56." In the shown orientation the N-terminus is on top. Image reproduced by the VMD software.[12]

COMPUTATIONAL METHOD

Theoretical background

Our approach basically does not differ from that of Yang and Honig.[13] Here we present the differences in the strategy chosen for the calculations. The electrostatic term of the free energy of unfolding at given pH can then be obtained (see Yang and Honig[13] and the references therein) by

$$\Delta G^{el}(\mathrm{pH}) = \Delta G^{el,U}(\mathrm{pH}) - \Delta G^{el,F}(\mathrm{pH})$$

$$= 2.3RT\int_{\mathrm{pH}_0}^{\mathrm{pH}}\Big(\nu^{U}(\mathrm{pH}) - \nu^{F}(\mathrm{pH})\Big)\mathrm{dpH}, \qquad (1)$$

where

$$\nu^{U,F}(\mathrm{pH}) = \sum_{i=1}^{N}\theta_i^{U,F}(\mathrm{pH}) \qquad (2)$$

are the average number of protons bound to the protein molecule in the unfolded (U) and in the folded (F) state, respectively, θ_ι is the degree of protonation of titratable site *i*, and *N* is the total number of titratable sites. $\Delta G^{el,U}$ and $\Delta G^{el,F}$ are the electrostatic free energies of the unfolded and folded states, respectively, and are defined in respect to a certain reference state. Different reference states can be chosen. For instance, Yang and Honig[13] have chosen pH equal to 0, whereas Langella *et al.*[14] defined it as an abstract state, at which all titratable groups in the protein are in their neutral forms. In this work, we choose another abstract reference state; that is an extreme acidic pH, pH_0, at which the protein is fully protonated in both, the folded and unfolded states. Obviously, $\theta_\iota^{U}(\mathrm{pH}_0) = \theta_\iota^{F}(\mathrm{pH}_0)$. Following Bashford and Karplus[15,16] the degree of protonation of site *i* is given by:

$$\theta_i(\mathrm{pH}) = \frac{\sum_{\{\mathbf{x}\}} x_i e^{-E(\mathbf{x},\mathrm{pH})/RT}}{\sum_{\{\mathbf{x}\}} e^{-E(\mathbf{x},\mathrm{pH})/RT}}. \qquad (3)$$

The energy $E(\mathbf{x},\mathrm{pH})$ is the electrostatic energy corresponding to a single protonation state, $\mathbf{x}$, of the protein molecule:

$$E(\mathbf{x},\mathrm{pH}) = 2.3RT\sum_{i}^{N} x_i(\mathrm{p}K_{i,\mathrm{int}} - \mathrm{pH}) + \frac{1}{2}\sum_{\substack{i,j\\ i\neq j}} W_{ix_i,jx_j}. \qquad (4)$$

Table I

Values of pK$_{mod}$ and pK of Lpp-56 Calculated for I = 0.12

Group	pK_{mod}[a]	*A*	*B*	*C*
N-term	8.2[b]	8.0	7.2	8.0
Asp07	4.0	2.0	2.8	1.9
Asp12		3.5	3.4	3.5
Asp21		3.3	3.2	3.2
Asp26		−1.9	−4.8	−3.7
Asp33		0.8	1.8	1.5
Asp39		2.6	3.5	0.2
Asp40		2.4	1.8	3.8
Asp49		2.5	2.6	0.8
Tyr55	9.4	16.1	15.9	16.0
Lys05	10.4	10.3	10.3	10.1
Lys19		10.6	10.2	10.7
Lys38		13.9	13.1	13.1
Lys54		13.0	12.8	13.2
Arg31	12.0	16.6	13.9	16.2
Arg43		14.4	12.9	14.0
Arg47		13.9	16.5	14.1
Arg56		12.5	12.6	12.5
C-ter	3.6[c]	−3.8	−2.9	−3.0

The individual helices are designated by *A*, *B*, and *C*.

[a] Values taken from Refs. 16, 17.

[b] p*K* of glycine amide (science.smith.edu/departments/Biochem/Biochem_353/Common_Buffers.htm).

[c] p*K* of *N*-acetyl glycine.[18]

A single protonation state is determined by the sequence $\mathbf{x} = (x_1, x_2, \ldots, x_i, x_j, \ldots, x_N)$, the elements of which describe the microscopic protonation states of the individual titratable sites. In this work, we assume that the individual titratable sites have only two microscopic states: $x_i = 0$ (protonated) and $x_i = 1$ (deprotonated). The electrostatic energy of interaction between sites *i* and *j* in protonation state x_i and x_j, respectively, is given by W_{ix_i,jx_j}, whereas $\mathrm{p}K_{i,\mathrm{int}}$ is the intrinsic p*K* value of site *i* defined as

$$\mathrm{p}K_{\mathrm{int}} = \mathrm{p}K_{i,\mathrm{mod}} + \Delta\mathrm{p}K_{i,\mathrm{Born}} + \Delta\mathrm{p}K_{i,\mathrm{pc}}. \qquad (5)$$

The first term in the right hand side of the above equation, $\mathrm{p}K_{i,\mathrm{mod}}$, is the equilibrium constant of a model compound (Table I). The correction $\Delta\mathrm{p}K_{i,\mathrm{Born}}$ accounts for the desolvation of site *i*, whereas $\Delta\mathrm{p}K_{i,\mathrm{pc}}$ is the p*K* shift caused by the electrostatic interactions of this site with the partial atomic charges of the protein molecule that do not belong to any titratable group.

Equations (3) and (4) are valid for both folded and unfolded states. The calculations, however, are based on different models. In the case of folded proteins the values of W_{ix_i,jx_j}, $\Delta\mathrm{p}K_{i,\mathrm{Born}}$ and $\Delta\mathrm{p}K_{i,\mathrm{pc}}$ are calculated on the basis of the three-dimensional structure of the protein. Prediction of the electrostatic properties of proteins based on crystal structures often faces difficulties arising from the fixed three-dimensional structure itself. For instance, due to the effect of the crystal contacts, regions of the protein molecule may preferably adopt conformations that are not populated in solution. In general, the X-ray structure does not necessarily represent the ensemble of structures of the protein in solution. This property of the protein structure, which we refer to as conformational flexibility, is one of the sources of discrepancy between prediction and experiment. Variety of methods exist that account in different extent for the conformational flexibility.[13,19–25]

The combination of molecular dynamics (MD) and pK calculations is an alternative approach.[13,26–30] The combination of MD and pK calculations results in an overall improvement of the theoretically predicted pK values. However, discrepancies between experimental and calculated pK values often remain for groups buried in the protein interior. Among different reasons one can mention the limited simulation time, as well as the bias of the simulation by the choice of the protonation state of the protein. Nevertheless, in our opinion, this method is currently the most promising one. In the investigation presented in this article, the ionization equilibria in the folded state of Lpp-56 have been calculated by a continuum electrostatic model combined with MD simulation as described in our previous work.[31]

The titration curves of unfolded proteins, ν^U(pH), are often calculated on the basis of the standard ionisation constants of the different types of titratable groups. In this null approximation electrostatic interactions are in fact ignored. This approximation is insufficient for prediction of quantities, such as the electrostatic term of unfolding energy.[32] The models of denatured state of proteins use different approximation mainly reflecting the increased hydration of the titratable groups.[33–36] A more general model of denatured state has been proposed by Zhou.[37–40] In this model, the denatured protein molecule is treated as a Gaussian chain immersed in a dielectric medium for which electrostatic interactions are calculated by means of the Debye-Hückel theory. Recently, we have proposed an approach[41–43] which is based on the continuum dielectric model and is ideologically very close to that of Zhou. In our model, the unfolded protein molecule is represented as a material with low dielectric constant, ε_p between 20 and 40, immersed in the high permittivity medium of the solvent, $\varepsilon_s > \varepsilon_p$. The shape of the dielectric cavity can be considered as an average over all possible conformations of the polypeptide chain, which results in a sphere inside which most of the protein atoms reside.[41] At equilibrium, the titratable groups are approximated as charge points allocated on the surface of the sphere. The charge distributions corresponding to the variety of conformers which an unfolded protein can adopt is reflected by the different configurations of a virtual chain connecting consecutively the charges on the surface of the dielectric cavity.[41] The model has been successfully applied for calculation of the pH-dependence of the unfolding free energy of several proteins.[44] This model has been employed in the presented study.

Molecular dynamics simulation

The X-ray structures of Lpp-56 (PDB entry 1eq7) and GCN4 leucine zipper (PDB entry 2zta) were used as starting point for MD simulations. The simulations were carried out with the OPLS all-atom force field, as implemented in the GROMACS simulation suite (version 3.3.1).[45] The structure was solvated with TIP4 water[46] at ~150 mM NaCl (plus additional ions to neutralize the total system). In the cubic periodic box the minimum distance between the protein and the end of the box was more 1.5 nm. After minimization using Steepest Descent model with a tolerance of 1000 kJ mol^{-1} nm^{-1}, the system was simulated for 400 ps with harmonic position restrain on all Cα-atoms (force constant: 1000 kJ mol^{-1} nm^{-2} and additional 100 ps with a force constant of 100 kJ mol^{-1} nm^{-2}) in order to allow relaxation of the solvent molecules. LINCS[47] and SETTLE[48] algorithms were applied. The integration step was 2 fs. Short-range electrostatics were calculated explicitly, and long range electrostatic interactions were calculated using the particle-mesh Ewald method.[49] Lennard-Jones interactions were cut at a distance of 1 nm, a long-range correction for the energy and the pressure was applied. The system was coupled to Berendsen temperature bath separately for the protein and the solvent (τ_t = 0.1 ps) and to a Berendsen pressure bath (τ_p = 0.1 ps).[50] Trajectory visualization and analyzing were made in part using VMD.[12]

p*K* calculations

The computational approach used to calculate θ_i^F(pH), and respectively ν^F(pH$_0$), has been described earlier.[31,51] The values of pK_{mod} are listed in Table I. The values of $\Delta pK_{i,Born}$ and $\Delta pK_{i,pc}$, required to complete Eq. (5), as well as the values of W_{ix_i,jx_j} in Eq. (4) were calculated by solving the linearized Poisson-Boltzmann equation using the finite difference method.[52,53] A programme developed in our laboratory was used for the calculations. The following parameters were used in the calculations: CHARMM parameter set 22[54] partial atomic charges, including those of the titratable groups in protonated and deprotonated states; van der Waals radii taken from Rashin *et al.*[55]; solvent probe radius 1.4 Å; ion exclusion layer of 2 Å. Each snapshot structure was situated in a grid box (99 × 99 × 99) with grid spacing of 2.55 Å, which was gradually reduced using four consecutive focusing steps on each titratable group. The size of the focused boxes depends on the conformations of the titratable side chains in the different snapshot structures. On average, the final grid length was 0.24 Å. Solvent and protein relative dielectric constants were taken ε_s = 78 and ε_p = 4, respectively.

The pK calculations for denatured state were performed according the procedure described in our earlier articles.[41,44] A radius of 17.5 Å and dielectric constant of 25[44] were used for the low dielectric sphere representing the unfolded protein.

Coupling p*K* calculations with MD simulation

Protein conformers (snapshot structures) were collected each 5 ps during the last 7 ns of the MD simulation. The calculations of electrostatic interactions were performed independently for each individual snapshot structure. The degree of deprotonation of the individual titratable sites, $\overline{\theta}_i$(pH), used in Eq. (2) (and then for calculation of ΔG^{el}) represent an arithmetic average of θ_i(pH) calculated for the individual snapshot structures. Test calculations showed that averaging over snapshot structures extracted in 5 ps and each 10 ps interval gives practically identical results. The results presented below are obtained by averaging over the interval of 10 ps.

RESULTS AND DISCUSSION

Starting from the X-ray structure we generated an ensemble of Lpp-56 conformers by a 10 ns MD simulation in explicit water. According to the usual criteria (C_α RMSD, radius of gyration, solute–solute and solute–solvent energy terms, intermolecular distances) the MD trajectory was well equilibrated in the last 7 ns of simulation, as to serve as a reliable model of the dynamic behavior of the protein. We first describe the results on prediction of the ionization properties of all Lpp-56 titratable groups. In a second part, we attempt quantification of the electrostatic contribution to Lpp-56 stability. Finally, we discuss a simple model providing clues about networked salt bridges as a possible player in the extremely slow unfolding transition of Lpp-56.

Ionisation equilibria

Table II
Comparison of the p*K Values of the Titratable Groups of Lpp-56* (I = 0.12) *Calculated on the Basis of the X-ray Structure* ($pK_{X\text{-}ray}$)*, Averaged Over the Structures Collected by MD Simulation* (pK_{MD}) *and of Unfolded State of the Protein* (pK_U)

Group	$pK_{X\text{-ray}}$	pK_{MD}	pK_U
N-term	9.5	7.8	8.6
Asp07	1.6	2.2	3.6
Asp12	5.3	3.5	3.7
Asp21	2.2	3.3	3.6
Asp26	2.2	−3.4	3.7
Asp33	3.6	1.4	3.8
Asp39	3.4	2.1	3.6
Asp40	0.8	2.7	3.6
Asp49	1.2	2.0	3.5
Tyr55	13.3	16.0	9.5
Lys05	10.1	10.2	10.8
Lys19	11.0	10.5	10.9
Lys38	8.8	13.4	10.9
Lys54	10.4	13.0	10.8
Arg31	12.2	15.6	12.7
Arg43	13.7	13.8	12.7
Arg47	16.1	14.8	12.7
Arg56	113.0	12.5	12.6
C-term	3.8	−3.2	2.9

The calculated p*K* values of the titratable sites of Lpp-56 are listed in Table I. As already mentioned, we are not aware of any experimental data that could help assessing the reliability of this result. According to our experience with other proteins, for instance ribonuclease T_1,[56] we expect that the confidence interval of calculated p*K* values is about 0.5 pH units. Inspecting Table I one can notice that groups situated at equivalent positions along the sequence of the three peptide chains have different p*K* values, which can deviate from each other by as much as 3 pH units. This reflects the fact that the corresponding side chains within the three helices visit different sets of conformations, thus creating different time-averaged environment of the titratable sites. In contrast, due to the threefold noncrystallographic symmetry, the p*K* calculated for equivalent sites using the X-ray structure vary within the confidence interval. Hence, the large p*K* deviation is not due to the computational method. Furthermore, about half of the titratable sites display p*K* values that undergo larger shifts from their standard (pK_{mod}) values when the calculations are done by averaging over the snapshot structures, in comparison to p*K* shifts calculated with the X-ray structure (Table II). The extremely up-shifted p*K* of basic groups and extremely down-shifted p*K* of acidic groups reflect a strong favorable electrostatic influence of the environment, such as participation in salt bridges. For a MD-generated ensemble of structures, the magnitude of the p*K* shift of groups participating in salt bridge depends on the population of conformers bearing the particular salt bridge. The effect of reduction of the p*K* shift because of the temporary disruption of a salt bridge which is illustrated in Figure 2 on the example of the salt bridge formed between the C-terminal carboxyl group (chain C) and LysB54. (Hereafter, the polypeptide chain to which a particular residue belongs is indicated by a capital letter inserted between the side-chain name and the sequence number.) In the time window in which no salt bridge is formed, the average p*K* of the C-terminal carboxyl group of chain C is practically equal to pK_{mod}. A shift of the p*K* value of more than 6 pH units occurs if the salt bridge with LysB54 is formed. Considering the first 2.5 ns of the simulation, the calculated average p*K* is approximately in the middle between the values corresponding to free and salt bridged C-terminal carboxyl group, reflecting the fact that the fractional populations of the free and salt-bridged C-terminal carboxylate are also approximately equal. For the rest of the simulation after 2.5 ns, however, the population of the salt bridge is virtually 100% (that is, it does not break), and consequently, the average p*K* of the monitored C-terminal carboxyl group continuously reduces. The groups with extreme p*K* shifts, such as the aspartic acids at position 26 or the considered above C-terminal groups, participate in salt bridges which seldom disrupt during the MD simulation.

Short lifetime (between 300 and 500 ps) of salt bridges suggested by some previous MD simulations[31,57] cannot

DOI 10.1002/prot

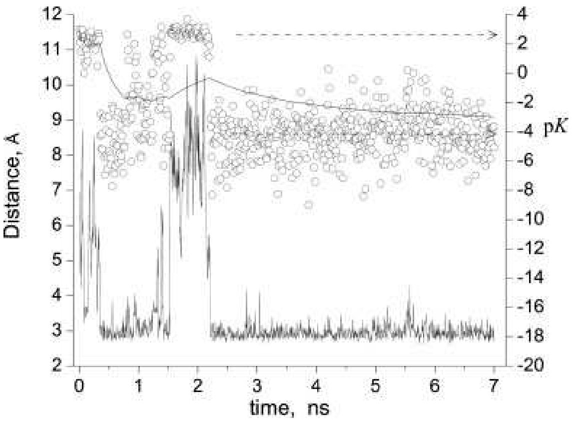


Figure 2

Salt bridge formation between the C-terminal carboxyl group (chain C) and Lys54 (chain B). Left ordinate: distances between the carboxyl oxygen atoms, (Oδ1+Oδ2)/2, and Nε of LysB54 (line). Right ordinate: snapshot pK values of the C-terminal carboxyl group (○), and time evolution of the average pK value of the C-terminal carboxyl group (line associated with the circles). The dashed arrows indicate the average pK when the carboxyl group does not form a salt bridge (upper arrow) and when salt bridge is formed (lower arrow).

be considered as a general rule. Salt bridges with lifetimes between 1 and 200 ns have been reported.[58–60] In principle, long lifetimes might be an artifact of the force field, if the oppositely charged atoms of the functional groups are trapped within short distances. To check this scenario, we performed MD simulation of the dimeric coiled-coil GCN4 using identical simulation protocol. This computational experiment revealed a different behavior of the titratable side chains involved in salt bridges: the lifetime of the salt bridges in GCN4 is essentially lower than that obtained for Lpp-56 (See Table IV). We conclude, therefore, that the long lifetime of some salt bridges and the extreme pK shifts calculated for Lpp-56 groups arise from the specific structural organization of the protein, rather than from a computational artifact.

The pK values calculated for the unfolded state of Lpp-56 are listed in Table II. All of them are non-negligibly shifted from their standard values. The average shift towards pK_{mod} of aspartic acids is 0.4 pH units, which is in accord with the experimental observation for other proteins.[61,62] This result reiterates the arising consensus that the denatured state is not an electrostatic "dummy" and that residual electrostatic effects in that state might contribute to the energetic balance stabilizing proteins.[63,64]

Electrostatic stabilization of Lpp-56

The electrostatic terms of the free energy calculated on the basis of the MD snapshot structures, $\Delta G^{\mathrm{el}}_{\mathrm{MD}}$, and using the X-ray structure only, $\Delta G^{\mathrm{el}}_{\mathrm{X}}$, are compared in Figure 3. The stabilizing contribution of the electrostatic interactions is substantially larger when calculated with the snapshot structures. This is an illustration of the effect of the reduced pK shifts calculated on the basis of the X-ray structure (Table II). The absolute values of ΔG^{el} have a meaning only in respect to the reference state, which here is an extreme low pH at which the folded and the unfolded states of the protein are identically protonated. Other values of ΔG^{el} will be obtained, as seen in Figure 3, if we choose the corresponding extreme alkaline pH. Therefore, no experimental verification of the absolute values of ΔG^{el} presented in Figure 3 can be made. However, if we assume that only electrostatic interactions change upon the change of pH, the relative pH dependence of ΔG^{el} can be verified experimentally. The unfolding free energy of Lpp-56 at pH 3 and pH 7 has been reported by Dragan *et al.*[5] From their data, $\Delta\Delta G^{\mathrm{therm}}(\mathrm{pH}\ 3 \rightarrow \mathrm{pH}\ 7) = \Delta G^{\mathrm{therm}}(\mathrm{pH}\ 7) - \Delta G^{\mathrm{therm}}(\mathrm{pH}\ 3)$ is ~60 kJ mol^{-1}. This value is close to $\Delta\Delta G^{\mathrm{el}}_{\mathrm{MD}}$ (pH 3 → pH 7) = 63 kJ mol^{-1} obtained by us using MD snapshot structures averaging. In contrast, calculations done with the X-ray structure predict a much smaller free energy change between pH 3 and pH 7: $\Delta\Delta G^{\mathrm{el}}_{\mathrm{X}}$ (pH 3 → pH 7) = 29 kJ mol^{-1}. The presented results are clear evidence that the introduction of conformational flexibility in the calculations of electrostatic interactions in protein improves the predictive power of the computations.

From equilibrium and kinetic data collected in the presence of GdmCl as the denaturant we recently estimated the stability of Lpp-56 as $\Delta G^{\mathrm{Gdm}} = 79 \pm 10$ kJ mol^{-1} pH 7.[4] At the same conditions, thermal unfolding experiments have predicted much higher stability,

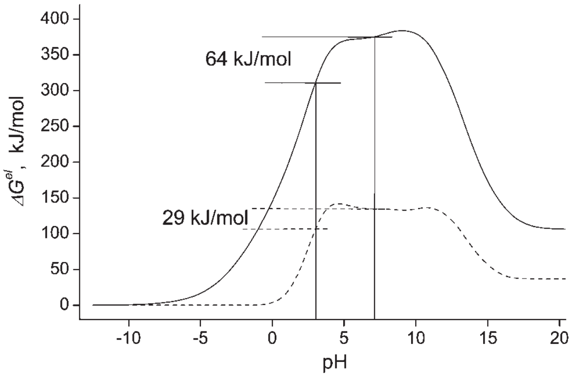


Figure 3

The electrostatic free energy as a function of pH calculated on the basis of MD simulation (continuous line) and the X-ray structure (dashed line).

$\Delta G^{\text{therm}} = 137$ kJ mol^{-1}.[5] It is commonly appreciated that GdmCl screens charge–charge interactions, so that the electrostatic contribution, ΔG^{el}, to ΔG^{Gdm} becomes smaller as the GdmCl concentration increases. Hence, the difference between ΔG^{Gdm} and ΔG^{therm} should be essentially electrostatic in nature.[6] ($\Delta G^{\text{Gdm}} < \Delta G^{\text{therm}}$ indicate that electrostatic effects are stabilizing the protein.) Taking into account the large, experimentally observed difference between ΔG^{Gdm} and ΔG^{therm} one can presume that the contribution of electrostatic interactions to the stability of the Lpp-56 is much larger than that observed for other proteins (typically less than 20 kJ mol^{-1}). Indeed, we have reasoned[4] that the linear extrapolation of the unfolding free energy from high GdmCl to zero denaturant to obtain ΔG^{Gdm} is very unlikely to underestimate the genuine stability of Lpp-56 by more than 20 kJ mol^{-1}. Nonetheless, the reasons for the large discrepancy between ΔG^{Gdm} and ΔG^{therm} remain obscure and, in fact, cannot be discerned by experiment. In the following, we discuss a computational approach to the problem.

It is believed that the energetic contribution of charge–charge interactions vanish around 1 *M* GdmCl. This statement, however, cannot be rigorously justified if one considers the influence of GdmCl of protein stability as a salt effect. The effect of ionic strength depends on the charge content[65,66] and on the charge configuration in the native state.[67] This suggests nonuniformed shielding of charge–charge interactions in folded and unfolded states.[68] Hence, the influence of salt concentration on stability is different for different proteins and no general statement about its magnitude can be done. We are interested in the influence of GdmCl on the magnitude of the electrostatic stabilization of Lpp-56. For concentrations of GdmCl less than 1*M* the reduction (or strengthening) of electrostatic stabilization can be considered as an effect of the ionic strength. In this way, the screening effect of GdmCl is reduced to calculations of ΔG^{el} for different ionic strengths:

$$\Delta^{\text{S}}\Delta G^{\text{el}} = \Delta G^{\text{el,S}} - \Delta G^{\text{el}} \tag{6}$$

Superscript S indicates high salt concentration, $I = 1M$. The calculated pH dependence of $\Delta^{\text{S}}\Delta G^{\text{el}}$ is shown in Figure 4. Because ΔG^{el} and $\Delta G^{\text{el,S}}$ are defined up to additive constants, their values at the reference state (chosen to be at an extremely acidic pH) are set to zero and $\Delta^{\text{S}}\Delta G^{\text{el}}_{\text{ref}} = 0$. In respect to this reference state at pH 7 $\Delta^{\text{S}}\Delta G^{\text{el}} = -14$ kJ mol^{-1}. However, the latter figure has no sound physical meaning (and for that matter can not be considered as representing the difference $\Delta G^{\text{therm}} - \Delta G^{\text{Gdm}}$) since the reference states in low and high ionic strengths are equalized.

The difference between the reference states used to calculate ΔG^{el} and $\Delta G^{\text{el,S}}$ can be evaluated. For this purpose, we make use of the thermodynamic cycle

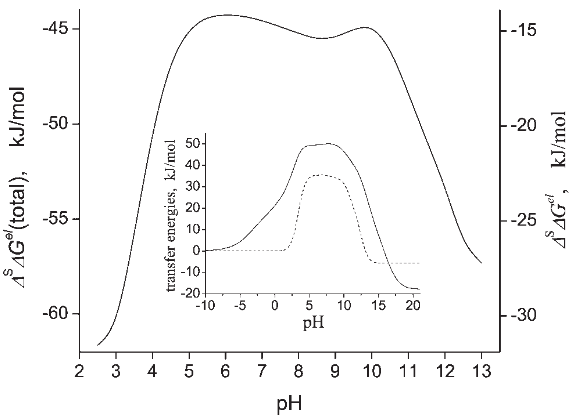


Figure 4

pH-Dependence of $\Delta^S\Delta G^{el}$ (right ordinate) and $\Delta^S\Delta G^{el}(pH)+\Delta G^{el}(F\rightarrow F^S)_{ref}$ (left ordinate). Insert: pH dependence of $\Delta G^{el}(F\rightarrow F^S)$ (continuous line) and $\Delta G^{el}(U\rightarrow U^S)$ (dashed line). The difference between these energies [Eq. (8)] gives $\Delta^S\Delta G^{el}$ (right ordinate).

$$\begin{tikzcd}
\mathrm{F} \arrow[r, "\Delta G^{\text{el}}"] \arrow[d, "\Delta G^{\text{el}}\ (\mathrm{F}\rightarrow\mathrm{F}^{\text{S}})"'] & \mathrm{U} \arrow[d, "\Delta G^{\text{el}}\ (\mathrm{U}\rightarrow\mathrm{U}^{\text{S}})"] \\
\mathrm{F}^{\text{S}} \arrow[r, "\Delta G^{\text{el,S}}"] & \mathrm{U}^{\text{S}}
\end{tikzcd}$$

The upper and lower horizontal limbs of the cycle represent unfolding at low ($I = 0.12M$) and high ($I = 1M$) ionic strengths, respectively. The left and right vertical limbs describe the hypothetical transfer of the folded (F) and unfolded (U) states, respectively, from low to high ionic strength conditions. According to the above thermodynamic cycle, the change of the electrostatic free energy of unfolding upon increase of the ionic strength given in Eq. (6) can also be expressed as:

$$\Delta^{\text{S}}\Delta G^{\text{el}} = \Delta G^{\text{el}}(\mathrm{U}\rightarrow\mathrm{U}^{\text{S}}) - \Delta G^{\text{el}}(\mathrm{F}\rightarrow\mathrm{F}^{\text{S}}). \tag{7}$$

Consider the reference state. It is chosen such that all titratable sites are protonated, that is, the protein contains positive charges only. According to our model of the unfolded state the charges tend to adopt positions, at which the repulsive forces, and hence electrostatic interactions, are minimized. In this aspect the model mimics well the real situation, where the denatured state is flexible and the charges could rearrange as to minimize the energetic penalty of charge–charge repulsion. Therefore, in a first approximation, we can assume that for the reference state $\Delta G^{\text{el}}(\mathrm{U}\rightarrow\mathrm{U}^{\text{S}})_{\text{ref}}$ is small and can be neglected. At pH far from the reference state $\Delta G^{\text{el}}(\mathrm{U}\rightarrow\mathrm{U}^{\text{S}})$ cannot be neglected, as illustrated in the insert of Figure 4. Analogous assumption for the reference state of the folded protein is not valid. The positions of the charges are fixed by the three-dimensional structure of the molecule, so that unfavorable electrostatic interactions between positive charges are sizeable in the reference state. Formally, the charge–charge interactions in the reference state can be calculated by Eq. (4) where the sequence $\mathbf{x} = (0, \ldots, 0)$ corresponds to all titratable sites in their protonated forms. The change of the charge–charge interactions upon the transfer of the native form of the protein in the reference state from low to high ionic strength calculated in this way amounts to -30 kJ mol^{-1} (the reference state is stabilized at high ionic strength because of the reduction of the repulsive interactions). Thus, the total change of the electrostatic free energy caused by the screening effect of GdmCl calculated for pH 7 becomes $\Delta^{S}\Delta G^{el}(\text{total}) = -45$ kJ mol^{-1}.

A certain underestimation of the value of $\Delta^{S}\Delta G^{el}(\text{total})$ is to be expected because the calculations at high ionic strength were preformed with the linearised Poisson-Boltzmann equation. The linearized and the nonlinear Poisson-Boltzmann equations give very similar results at least up to $I = 0.5M$, underestimating the reduction of electrostatic energy because of the salt effect by about 10%.[69] Even ignoring this underestimation, it is clear that the value of $\Delta^{S}\Delta G^{el}(\text{total})$ represents a significant energetic contribution. The estimated 45 kJ mol^{-1} unfolding free energy reduction stemming from charge screening by salt (GdmCl for that matter) exceeds our previous estimate (10–20 kJ mol^{-1}). Rather, it approaches the value (~60 kJ mol^{-1}) corresponding to the difference between ΔG^{Gdm} measured by us[70] and ΔG^{therm} obtained by Dragan *et al.*[5]

We would like to add a note of caution in interpreting the numerical value of the charge–charge contribution to the stability of Lpp-56, presumed to represent the total difference of experimentally measured unfolding free energies ($\Delta G^{\text{therm}} - \Delta G^{\text{Gdm}}$). It has been argued that both, the native and unfolded states are not fixed in their properties, depending on the physical agent used to shift the equilibrium between these states.[9] The calculations presented in this work consider only the ionic strength effect. The influence of the electrolyte type, including protein–ion binding effects, is ignored. This is in fact ignoring any denaturant-specific differences in the structure of the folded and the unfolded states. In spite of all these considerations, we conclude that indeed the difference between the unfolding free energies of Lpp-56 measured by the two different experimental approaches is to a large extent due to the screening effect of GdmCl.

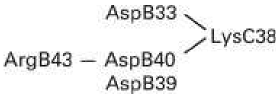
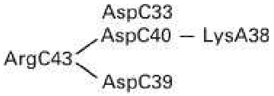
Role of the salt bridges in unfolding kinetic of Lpp-56

Although folding of Lpp-56 is also slow, it appears that the high thermodynamic stability originates from an extremely low unfolding rate. On the basis of the strong dependence of the unfolding rate constant on the concentration of salt (GdmCl) we have speculated that there is a significant electrostatic component to the activation energy barrier for unfolding. Here, we ask whether the known high kinetic stability of Lpp-56 can be related to this electrostatic component, in particular to the presence of salt bridges which seldom or never disrupt during the MD simulation.

We have mentioned that the trimeric superhelix is clamped along its length by rings of interhelical salt bridges (see Fig. 1). Since the prevailing majority of them have long lifetimes it is worth having a closer look at their behavior. In the course of the MD simulation we observe formation of 15 salt bridges. Among them 12 are interhelical links forming the charge rings illustrated in Figure 1. The organization of rings and the lifetimes of the salt bridges constituting them are specified in Table III. In the following we consider a salt bridge being

Table III

Interhelical Salt Bridges in Lpp-56

Ring 1	*p*
Asp21A—Lys19C	0.02
Asp21B—Lys19A	0.01
Asp21C—Lys19B	0
Ring 2	*p*
ArgA31—AspC26	1
ArgB31—AspA26	0.98
ArgC31—AspB26	0.99
Ring 3-cluster	*p*
	0.9
	0.79
	0.95
Ring 4	*p*
AspA49—ArgC47	0.52
AspB49—ArgA47	0.35
AspC49—ArgB47	0.92
Ring 5	*p*
CtrA56—LysC54	0.95
CtrB56—LysA54	0.89
CtrC56—LysB54	0.85
Ring 6	*p*
CtrA56—TyrC55	0.99
CtrB56—TyrA55	0.99
CtrC56—TyrB55	0.99

The parameter p is the ratio between the lifetime of a salt bridge and the total time (7 ns) of the MD simulation used to collect snapshot structures.

formed if at least one donor-acceptor distance between the bridged groups to be less than 3.1 Å. This distance corresponds to the upper limit for a stable hydrogen bond. Although electrostatic attraction between the interacting groups is significant even at distances larger than the chosen criterion such configurations lose the features of a hydrogen bond.

The group of salt bridges close to the N-terminus (top in Fig. 1) does not form a ring of interhelical rings. Of certain interest are the salt bridges forming ring 1. In the X-ray structure all these salt bridges are well defined, with proton donor-to-proton acceptor distances between 2.6 and 2.9 Å, corresponding to an ideal hydrogen bond. However, it turned out that the lifetime of these salt bridges is negligible during the MD simulation. Almost opposite is the situation in Ring 2, where the salt bridges practically do not disrupt during the MD simulation, whereas in the X-ray structure only one of them satisfies the above criterion. The difference in the behavior of these salt bridges is reflected by the large difference in the p*K* values of participating groups (see Table II for comparison).

Ring 3 is a cluster of salt bridges involving also intrahelical links. The pairs within the cluster exchange their partners as indicated in Table III. This feature is also illustrated in Figure 5 for the case of LysC38. This residue adopts conformations at which it preferably interacts with AspB33 or with AspB40. The lifetime of the salt bridges this side chain forms is plotted in Figure 6. It is important for our further considerations to note that the cross-link between the helices B and C is intact in spite of the mobility of the lysine side chain. Similar behavior is also observed for the intramolecular salt bridges with the participation of Arg43 (see Table III). This observation leads us to the conclusion that the stabilization role of this ring is achieved by both, favorable interhelical electrostatic interactions and reduction of entropic losses.

Ring 4 displays properties similar to those of Rings 2 and 5, yet it appears more "loose" judging from the average lifetime of the participating salt bridges (Table III).

Ring 5 is also stabilized by a network involving the hydrogen bond between C-terminal carboxyl groups and the hydroxyl groups of Tyr55. In contrast to the network of Ring 3, here the configurations HOη(TyrY55)–(LysX56)COO$^-$–NζH$_3^+$(LysX56) remain stable with lifetime of at least 85% (see Table III). As illustrated in Figure 7, the hydrogen bonds (Lys56)COO$^{-}\cdots$HOη(Tyr55) are interhelical and are expected to contribute to the stabilization of the bundle in this region. Because of this, we consider these hydrogen bonds formally as a separate ring (Ring 6).

The overview of the rings of interhelical salt bridges suggests that they should play an important role in the stabilization of the native three-dimensional structure of Lpp-56. On the basis of this, we hypothesize that the long life time of the prevailing majority of interhelical salt bridges contributes for structural stability of Lpp-56 as well as for its low unfolding rate.

Assume for simplicity that the protein unfolding is initiated by the disruption of the salt bridge rings. We consider a ring as broken if at least two interhelical salt bridges within this ring are simultaneously disrupted. This assumption reflects the properties of quasi-symmetric, noncovalent homotrimers, where dissociation of one monomer requires simultaneous disruption of two sets of quasi-symmetric interactions.[71] Two salt bridges being disrupted, one of the helices in the region of a given ring could more easily move away from the other two. The hydrophobic packing is weaken (leading to enthalpic destabilization), the mobility of groups increases (making the molecule more sensitive to thermal fluctuations), the hydrophobic core becomes partially hydrated. These effects promote non-native conformations. Such conformations may be stabilized and may propagate if the neighboring ring is broken, otherwise the native conformation is stabilized. Since unfolding is coupled to chain dissociation, according to this scenario, a successful attempt for unfolding occurs if all salt bridge rings break simultaneously. The probability a ring, *i*, to be disrupted can be calculated by

$$p_i = \overline{p_{AC}}\,\overline{p_{BA}}\,\overline{p_{CB}} + p_{AC}\overline{p_{BA}}\,\overline{p_{CB}} + \overline{p_{AC}}\,p_{BA}\overline{p_{CB}} + \overline{p_{AC}}\,\overline{p_{BA}}\,p_{CB}, \tag{8}$$

where p_{XY} is the probability a cross-link (an interhelical salt bridge) between helix X and helix Y to exist. The probability p_{XY} is the parameter p (given in Table III) calculated as the ratio between the lifetime of a salt bridge connecting helices X and Y and the total time of simulation. The probability a cross-link between helix X and helix Y to be disrupted is then $\overline{p_{XY}} = 1 - p_{XY}$. Equation (8) comprises of the sum of the probability all salt bridges in a ring to be disrupted (the first term on its right hand side) and the probabilities one salt bridge (p_{AC} or p_{BA} or p_{CB}) to be intact while the other two are disrupted. The equation is valid if the events XY are independent. Since we have not found any correlation between the breaking and formation of the salt bridges within the rings, the above condition can be considered as fulfilled. The probability of all rings are to be disrupted simultaneously, that is, the probability of a successful attempt for unfolding, is then

$$p_{\text{unf}} = \prod_{i=1}^{6} p_i = 2.0 \times 10^{-11}.$$

One can consider the process of unfolding as a series of events (independent or related), which stabilize non-native conformations. The unfolding rate will be then limited by the events with lower probability. In this context, we relate the extremely low value of the probability p_{unf} with the

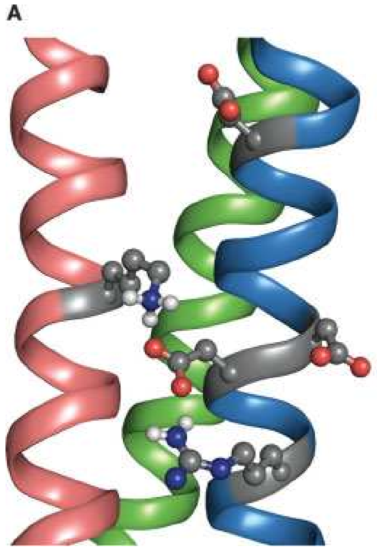


Figure 5

Part of the salt bridge network in Ring 3 (Table III). Snapshot structures ***A****: Salt bridge AspB33-LysC38. Snapshots structure* ***B****: Salt bridge AspB40-LysC38. Image reproduced using The PyMOL Executable Build, (2005) DeLano Scientific LLC, Sauth San Francisco, CA, USA.*

likewise low unfolding rate constant measured for Lpp-56. This correlation is valid if the time interval of the MD simulation is long enough to ensure ergodicity of the system. An evidence for this is in good agreement between the calculated change of the electrostatic free energy with pH, $\Delta\Delta G^{\mathrm{el}}_{\mathrm{MD}}$ (pH 3 → pH 7), and the experimentally observed $\Delta\Delta G^{\mathrm{therm}}$(pH 3 → pH 7). An additional indirect evidence is the value of p_{unf} calculated for the GCN4 leucine zipper. This protein has a spatial organization similar to that of Lpp-56, although it consists of two, instead of three α-helices. The lifetimes of the interhelical salt

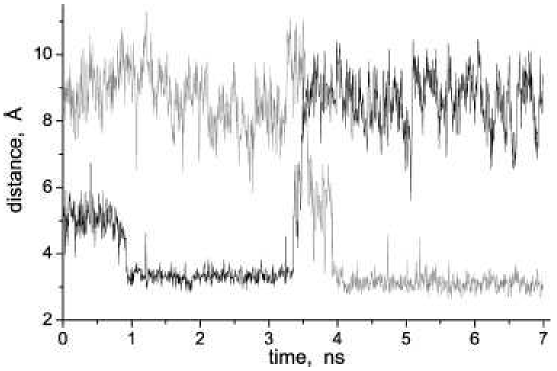


Figure 6

Lifetime of salt bridges AspB33-LysC38 (grey line) and AspB40-LysC38 (black line), The distance plotted is between the average coordinates of the carboxyl oxygen atoms, (Oδ1+Oδ2)/2, of the aspartic acid and the Nζ atom of LysC38.

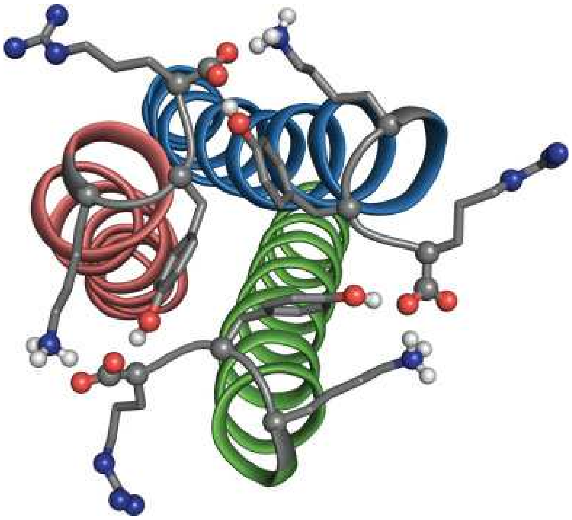


Figure 7

C-terminal Rings 5 and 6. Image reproduced using The PyMOL Executable Build, (2005) DeLano Scientific LLC, Sauth San Francisco, CA, USA.

Table IV

Interhelical Salt Bridges in GCN4 Leucine Zipper

Ring 1	*p*
LysA15—GluB20	0.48
LysB15—GluA20	0.30
Ring 2	*p*
GluA22—LysB27	0.12
GluB22—LysA27	0.45
Ring 3	*p*
ArgA25—CtrB31	0.07
ArgB25—CtrA31	0.10

The parameter p is described in the legend of Table III.

bridges in GCN4, grouped in rings by analogy with Lpp-56, are listed in Table IV. To maintain the assumptions as close as possible to those made for Lpp-56, a ring is considered as disrupted, if it consists of only one interhelical salt bridge. The probability of such an event is then

$$p_i = 1 - p_{AB}p_{BA}$$

and $p_{\text{unf}} = \Pi p_i = 0.79$. The probabilities p_{AB} and p_{BA} correspond to symmetric salt bridges connecting the two helices of GCN4. However the probabilities are different (Table IV) because of the independent side chains movement of the two helices. If we assume that a ring breaks when both salt bridges are disrupted

$$p_i = \overline{p_{AB}p_{BA}}$$

and $p_{\text{unf}} = 0.12$. The values of p_{unf} calculated for GCN4 are essentially larger than that calculated for Lpp-56. If we relate the probabilities for successful unfolding of the two proteins with their unfolding rate constants, it follows from the $p_{\text{unf}}^{\text{GCN4}}/p_{\text{unf}}^{\text{Lpp}}$ ratio that GCN4 unfolds 3.9×10^{10} to 5.9×10^{9} faster than Lpp-56. Unfortunately, a direct comparison with the actual unfolding constants is not possible, since the unfolding constant of Lpp-56 is not precisely known. However, it is $>1 \times 10^{-11}\ \text{s}^{-1}$. Since the unfolding rate constant of GCN4 is $\sim 2 \times 10^{-2}\ \text{s}^{-1}$,[72] the ratio of the experimentally measured rate constants is $k_{\text{unf}}^{\text{GCN4}}/k_{\text{unf}}^{\text{Lpp}} > 2 \times 10^{9}$. On the basis of the good agreement between our prediction and experiment observation, we conclude that the enormously low unfolding rate of Lpp-56 is essentially related to electrostatic interactions, and in particular, to the stability of the salt bridges.

It appears that the simple assumption underlying the model capture important features of Lpp-56 (and probably other coiled coils). The careful thermodynamic analysis of Dragan *et al.*[5] led the authors describe the Lpp-56 unfolding transition state as native-like, that is "... the helices forming this coiled coil are still not sufficiently separated ... yet at this stage extensive disruption of some short-range enthalpic interactions takes place...." Furthermore, they analyze the activation enthalpy and entropy of unfolding and conclude: "It appears thus that unfolding of the rigid three-stranded coiled coil starts from the simultaneous disruption of all van der Waals contacts between the strands, and since the probability of that is low, the process is slow." Such a picture is fully complementary to the basic assumptions of the presented model. It appears that the long-living salt bridges, which are staggered along the rod-like molecule, effectively prevent propagation of local unfolding events. Since it is believed that end-fraying is a factor destabilizing coiled coils, of special interest is the fact that the C-terminus of the molecule is tightly constrained through electrostatic interactions within Rings 5 and 6.

Finally, we would like to add that also in other proteins salt bridges which cross-link secondary structure elements, or are present at the interface of sub-units might provide a source of kinetic stabilization, possibly by reducing the activation entropy for unfolding, thus increasing the activation energy for unfolding.[60,73–75] It should be noted, however, that the stabilizing effect of the salt bridge charge–charge interactions is a consequence of the dynamic properties of the groups involved. These properties are, on the other hand, determined by the dynamic properties of the environment, which may or may not tolerate conformational freedom of the charged side chains, in this way regulating the salt bridge lifetime. In general, this feature cannot be recognized from a single, say crystal, protein structure. In this context, the combination of MD-based analysis and pK calculations might serve as a useful guide for experimentalists in mutation-based approaches.

ACKNOWLEDGMENTS

The molecular dynamics simulations were performed on the Matterhorn Beowulf cluster at the Informatikdienste of the University of Zurich. The authors thank C. Bollinger, Dr. T. Steenbock, and Dr. A. Godknecht for setting up and maintaining the cluster.

REFERENCES

1. Lupas AN, Gruber M. The structure of α-helical coiled coils. Adv Protein Chem 2005;70:37–78.
2. Shu W, Liu J, Ji H, Lu M. Core structure of the outer membrane lipoprotein from *Escherichia coli* at 1.9 Å resolution. J Mol Biol 2000;299:1101–1112.
3. DiRienzo JM, Nakamura K, Inouye M. The outer membrane proteins of gram-negative bacteria: biosynthesis, assembly, and functions. Annu Rev Biochem 1978;47:481–532.
4. Bjelic S, Karshikoff A, Jelesarov I. Stability and folding/unfolding kinetics of the homotrimeric coiled coil Lpp-56. Biochemistry 2006;45:8931–8939.
5. Dragan AI, Potekhin SA, Sivolob A, Lu M, Privalov PL. Kinetics and thermodynamics of the unfolding and refolding of the three-stranded a-helical coiled coil, Lpp-56. Biochemistry 2004;43:14891–14900.
6. Ibarra-Molero B, Loladze VV, Makhatadze GI, Sanchez-Ruiz JM. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge to interactions protein stability. Biochemistry 1999;38:8138–8149.

7. Makhatadze GI. Thermodynamics of protein interactions with urea and guanidinium hydrochloride. J Phys Chem B 1999;103:4781–4785.
8. Ibarra-Molero B, Sanchez-Ruiz JM. A model-independent, nonlinear extrapolation procedure for the characterization of protein folding energetics from solvent-denaturation data. Biochemistry 1996;35:14689–14702.
9. Ferreon AC, Bolen DW. Thermodynamics of denaturant-induced unfolding of a protein that exhibits variable two-state denaturation. Biochemistry 2004;43:13357–13369.
10. Pfeil W. Protein stability and folding: a collection of thermodynamic data. New York: Springer-Verlag; 2001.
11. Monera OD, Kay CM, Hodges RS. Protein denaturation with guanidine-hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. Protein Sci 1994;3:1984–1991.
12. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38.
13. Yang A-S, Honig B. On the pH dependence of protein stability. J Mol Biol 1993;231:459–474.
14. Langella E, Improta R, Crescenzi O, Barone V. Assessing the acid-base and conformational properties of histidine residues in human prion protein (125–228) by means of pK_a calculations and molecular dynamics simulations. Proteins 2006;64:167–177.
15. Bashford D, Karplus M. Multiple-site titration curves of proteins: an analysis of exact and approximate methods for their calculation. J Phys Chem 1991;95:9556–9561.
16. Bashford D, Karplus M. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. Biochemistry 1990;29:10219–10225.
17. Matthew JB. Electrostatic effects in proteins. Annu Rev Biophys Biomol Struct 1985;14:387–417.
18. Weast RC, editor. Dissociation constants of organic bases in aqueous solution. Handbook of chemistry and physics, 51st edition. Ohio: The Chemical Rubber CO.; 1971. p D117.
19. Yang A-S, Gunner MR, Sampogna R, Sharp K, Honig B. On the calculation of pK_as in proteins. Proteins 1993;15:252–265.
20. You TJ, Bashford D. Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. Biophys J 1995;69:1721–1733.
21. Beroza P, Case DA. Including side chain flexibility in continuum electrosttatic calculations of protein titration. J Phys Chem 1996;100:20156–20163.
22. Alexov E, Gunner MR. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. Biophys J 1997;72:2075–2093.
23. Alexov EG, Gunner MR. Calculated protein and protonmotion coupled to electron transfer: electron transfer from Q_A-Q_B to Q_B in bacterial photosynthetic reaction centers. Biochemistry 1999;38: 8253–8270.
24. Georgescu RE, Alexov E, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. Biophys J 2002;83:1731–1748.
25. Koumanov A, Benach J, Atrian S, Gonzàlez-Duarte R, Karshikoff A, Ladenstein R. The catalytic mechanism of Drosophila alcohol dehydrogenase: evidence for a proton relay modulated by the coupled ionization of the active site lysine/tyrosine pair and a NAD^+ ribose OH switch. Proteins 2003;51:289–298.
26. Bashford D, Gerwert K. Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. J Mol Biol 1992;224:473–486.
27. Yang A-S, Honig B. Structural origin of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. J Mol Biol 1994;237:602–614.
28. Zhou H-X, Vijayakumar M. Modeling of protein conformational fluctuations in pK_a prediction. J Mol Biol 1997;267:1002–1011.
29. van Vlijmen HWT, Schaefer M, Karplus M. Improving the accuracy of protein pK_a calculations—conformational averaging versus the average structure. Proteins 1998;33:145–158.
30. Baptista AM, Teixeira VH, Soares CM. Constant-pH molecular dynamics using stochastic titration. J Chem Phys 2002;117:4184–4200.
31. Koumanov A, Karshikoff A, Friis EP, Borchert TV. Conformational averaging in pK calculations. Improvement and limitations in prediction of ionization properties of proteins. J Phys Chem B 2001;105:9339–9344.
32. Pace CN, Alston RW, Shaw KL. Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. Protein Sci 2000;9:1395–1398.
33. Schaefer M, Sommer M, Karplus M. pH-dependence of protein stability: absolute electrostatic free energy difference between conformations. J Phys Chem B 1997;101:1663–1683.
34. Warwicker J. Simplified methods for pK(a) and acid pH-dependent stability estimation in proteins: removing dielectric and counterion boundaries. Protein Sci 1999;8:418–425.
35. Phelan P, Gorfe AA, Jelesarov I, Marti DN, Warwicker J, Bosshard HR. Salt bridges destabilize a leucine zipper designed for maximized ion pairing between helices. Biochemistry 2002;41:2998–3008.
36. Elcock AH. Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability. J Mol Biol 1999;294:1051–1062.
37. Zhou HX. A Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins. Proc Natl Acad Sci USA 2002;99:3569–3574.
38. Zhou HX. Residual electrostatic effects in the unfolded state of the N-terminal domain of L9 can be attributed to nonspecific nonlocal charge-charge interactions. Biochemistry 2002;41:6533–6538.
39. Zhou HX. Dimensions of denatured protein chains from hydrodynamic data. J Phys Chem B 2002;106:5769–5775.
40. Zhou HX. Direct test of the gaussioan-chain model for treating residual charge-charge interactions in the unfolded state of proteins. J Am Chem Soc 2003;125:2060–2061.
41. Kundrotas PJ, Karshikoff A. Model for calculations of electrostatic interactions in unfolded proteins. Phys Rev E 2002;65(1 Part 1): 011901.
42. Kundrotas PJ, Karshikoff A. Effects of electrostatic interactions on dimensions of unfolded polypeptide chains with various charge distributions: monte carlo study. J Chem Phys 2003;119:3574–3581.
43. Kundrotas PJ, Karshikoff A. Charge sequence coding in statistical modeling of unfolded proteins. Biochim Biophys Acta 2004;1702:1–8.
44. Kundrotas PJ, Karshikoff A. Modeling of denatured state for calculations of electrostatic contribution to protein stability. Protein Sci 2002;11:1681–1686.
45. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 2001; 7:306–317.
46. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
47. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997; 18:1463–1472.
48. Miyamoto S, Kollman PA. Settle–an analytical version of the shake and rattle algorithm for rigid water models. J Comput Chem 1992;13:952–962.
49. Darden T, York D, Pedersen L. Particle mesh ewald—an N.log(N) method for ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
50. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR. Molecular-dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
51. Karshikoff A. A simple algorithm for calculation of multiple site titration curves. Protein Eng 1995;8:243–248.
52. Warwicker J, Watson NC. Calculation of the electric field potential in the active site cleft due to α-helix dipoles. J Mol Biol 1982;157: 671–679.

53. Nicholls A, Honig B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. J Comput Chem 1991;12:435–445.
54. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983; 4:187–217.
55. Rashin AA, Iofin M, Honig B. Internal cavities and buried waters in globular proteins. Biochemistry 1986;25:3619–3625.
56. Koumanov A, Spitzner N, Rüterjans H, Karshikoff A. Ionisation properties of titratable groups in ribonuclease T_1. II. Electrostatic analysis Eur Biophys J 2001;30:198–206.
57. Caflisch A, Karplus M. Acid and thermal denaturation of barnase investigated by molecular dynamics simulation. J Mol Biol 1995;252:627–708.
58. Gruia AD, Fischer S, Smith JC. Kinetics of breaking a salt-bridge critical in protein unfolding. Chem Phys Lett 2004;385:337–340.
59. Sheldahl C, Harvey SC. Molecular dynamics on a model for nascent high-density lipoprotein: role of salt bridges. Biophys J 1999;76: 1190–1198.
60. Huang XQ, Zhou HX. Similarity and difference in the unfolding of thermophilic and mesophilic cold shock proteins studied by molecular dynamics simulations. Biophys J 2006;91:2451–2463.
61. Tan Y-J, Oliveberg M, Davis B, Fersht AR. Perturbed pK_a-values in the denatured states of proteins. J Mol Biol 1995;254:980–992.
62. Tollinger M, Forman-Kay JD, Kay LE. Measurement of side-chain carboxyl pK_a values of glutamate and aspartate residues in an unfolded protein by multinuclear NMR spectroscopy. J Am Chem Soc 2002;124:5714–5717.
63. Whitten ST, Garcia-Moreno B. pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. Biochemistry 2000;39:14292–14304.
64. Cho J-H, Raleigh DP. Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. J Mol Biol 2005;353:174–185.
65. Zhou H-X, Dong F. Electrostatic contributions to the stability of a thermophilic cold shock protein. Biophys J 2003;84:2216–2222.
66. Spencer DS, Xu K, Logan TM, Zhou HX. Effects of ph, salt, and macromolecular crowding on the stability of fk506-binding protein: an integrated experimental and theoretical study. J Mol Biol 2005; 351:219–232.
67. Lee KK, Fitch CA, Lecomte JTJ, García-Moreno BE. Electrostatic effects in highly charged proteins: salt sensitivity of pK_a values of histidines in staphylococcal. Biochemistry 2002;82: 5656–5667.
68. Lindman S, Xue WF, Szczepankiewicz O, Bauer MC, Nilsson H, Linse S. Salting the charged surface: pH and salt dependence of protein G B1 stability. Biophys J 2006;90:2911–2921.
69. Rocchia W, Alexov E, Honig B. Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. J Phys Chem 2001;105:6507–6514.
70. Karshikoff A, Ladenstein R. The role of electrostatic interactions in the stabilization of proteins from thermophiles. In: Uversky VN, Permyakov EA, editors. Protein structures: methods in protein structure and stability analysis. New York: Nova Science Publishers; 2006.
71. Forrer P, Chang C, Ott D, Wlodawer A, Plückthun A. Kinetic stability and crystal structure of the viral capsid protein SHP. J Mol Biol 2004;344:179–193.
72. Ibarra-Molero B, Makhatadze GI, Matthews CR. Mapping the energy surface for the folding reaction of the coiled-coil peptide GCN4-p1. Biochemistry 2001;40:719–731.
73. Cavagnero S, Debe DA, Zhou ZH, Adams MWW, Chan SI. Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins. Biochemistry 1998;37:3369–3376.
74. Bellapadrona G, Chiaraluce R, Consalvi V, Ilari A, Stefanini S, Chiancone E. The mutations Lys 114 to Gln and Asp 126 to Asn disrupt an intersubunit salt bridge and convert *Listeria innocua* DPS into its natural mutant *Listeria monocytogenes* Dps. Effects on protein stability at low pH. Proteins 2007;66:975–983.
75. Zubillaga RA, Garcia-Hernandez E, Camarillo-Cadena M, Leon M, Polaina J. Effect of a new ionic pair on the unfolding activation barrier of β-glucosidase B. Protein Pept Lett 2006;13:113–118.

CHAPTER 3

BIOPHYSICAL CHARACTERIZATION OF THE C-TERMINAL DOMAIN OF END-BINDING PROTEIN 1 (EB1C)

3.1 INTRODUCTION

The intrinsic dynamic properties of microtubules (MTs) are important in establishing and maintaining the specific organization of cellular components during the cell cycle. Diverse factors regulate MT dynamics, both spatially and temporally. Among these, a number of proteins and protein complexes have been identified that specifically track growing MT plus ends (+TIP proteins; [1]). End-binding protein 1 (EB1) belongs to a highly conserved and ubiquitously expressed family of +TIP proteins. The EB1 proteins are composed of three domains ([2] and Fig. 3.1A). The N-terminal domain exhibits a calponin homology fold and is responsible for recognition of the MT tip [3]. The structure of the intermediate domain is not known but this part of the molecule is probably very flexible and essentially unstructured as predicted from the amino acid composition and limited proteolysis results. Sequentially, the C-terminal domain (EB1c; residues 191-268) can be sub-divided into two sub-domains. The N-terminal sub-domain (Ca; residues 191-225) exhibits the characteristic 4,3 hydrophobic (heptad) repeat, indicating high probability for adopting an α-helical coiled coil conformation, while the C-terminal sub-domain (Cb, residues 226-268) has a low complexity sequence. The structure of EB1c domain was recently solved by X-ray crystallography at high resolution ([3] and Fig. 3.1B). EB1c is a dimer composed of two highly helical monomers arranged in parallel, in-register orientation. The N-terminal 40 residues long α-helices (referred to as $\alpha 1$ helices) wrap around each other to form a left-handed coiled coil. The coiled coil structure is maintained by the typical knobs-into-hole packing of residues occupying ***a*** and ***d*** heptad positions. The C-terminal ends of the $\alpha 1$ helices diverge beyond position 221 and extend into a non-helical hairpin segment, which is followed by short helices ($\alpha 2$ helices; residues 237-248). The $\alpha 2$ helix of each monomer packs into antiparallel orientation with the C-terminal part of the $\alpha 1$ helix of the adjacent monomer. Altogether the C-terminal quarter of EB1c

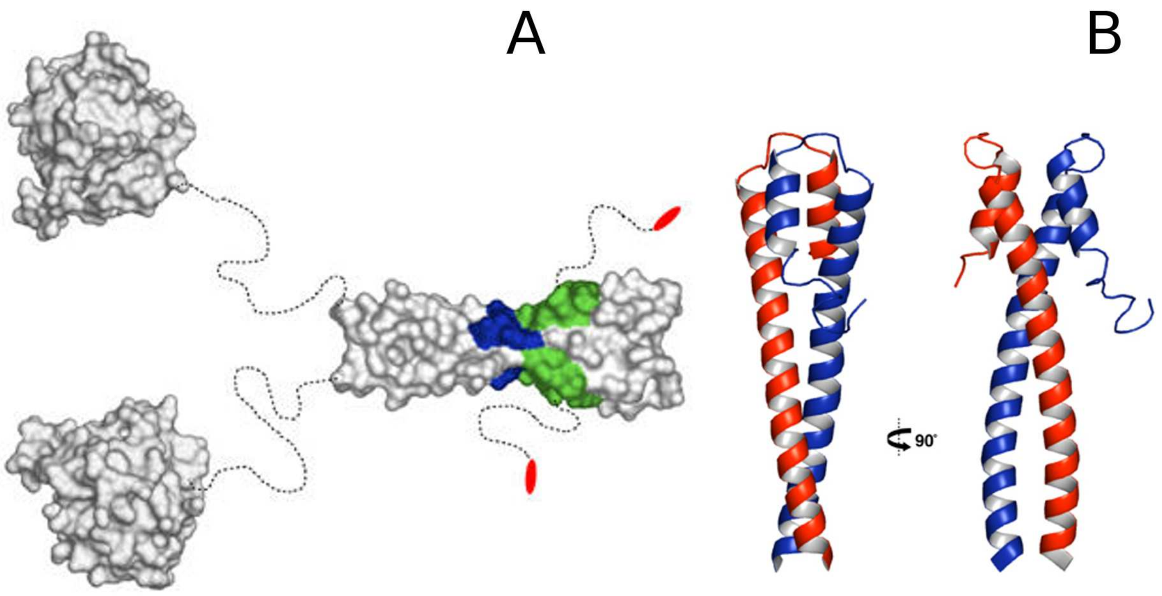


Fig. 3.1: Structural organization of EB1 proteins. (Panel A) Overall structure. The globular N-terminal, MT-binding calponin-homology domains (PDB entry 1pa7) and the C-terminal EB1-C domain (PDB entry 1wu9) are depicted as surface views. Dashed and curved lines schematize flexible peptide segments of the dimeric molecule. The binding sites for APC (hydrophobic cavity and polar rim) and dynactin /p150glued (hydrophobic cavity, polar rim and C-terminal EEY-COO$^-$ tripeptide) are highlighted in colors: green, hydrophobic cavity; blue, polar rim; red, C-terminal EEY-COO$^-$ tripeptide. (Panel B) EB1-C in cartoon representation. Monomers A and B are colored blue and red, respectively. (Adapted from ref. [2])

protein is represented by a four-helix bundle. There is no interpretable electron density for residues beyond residue 250, indicating that the very C-terminal amino acid stretch lacks structure and is flexible. It has been demonstrated that although the coiled coil subdomain Ca formally spans four heptads and is expected to have comparable stability to other four-heptad coiled coils, like the GCN4 leucine zipper, is in fact rather unstable in isolation and unfolds in a non-cooperative manner [2]. In contrast, EB1c undergoes a cooperative melting transition centered at around 65 ℃ (40 μM monomer concentration; [2]). These observations point to the role of the four-helix bundle to the overall stability of the protein. The aim of the experiments described bellow is to determine the thermodynamic parameters of EB1c stability, and to investigate the folding mechanism in terms of cooperativity and intermediary states by comparing calorimetric and effective van't Hoff parameters, and by comparing equilibrium and kinetic parameters[1].

---

[1] The experiments described in this chapter were done in collaboration with Oksana Okhrimenko.

3.2 EXPERIMENTAL PROCEDURES

Protein expression and purification. The construction of the EB1c expression vector can be found in [2]. The protein was expressed in *E. coli* strain BL21(DE3) (Novagen). The cells were grown at 37 ℃ in standard LB medium. After induction at OD(600)=0.6 by 1 mM IPTG the cells were grown at 37 ℃ for additional 5 hours or the temperature was reduced to 25 ℃ and the cells were incubated overnight. The cells were pelleted by centrifugation at 8000 g and frozen at -80 ℃. Cells were lyzed by french-press treatment under native conditions using lysis buffer (30 mM Tris-HCl, 500 mM NaCl, 6 mM imidazol, 1 mM beta-mercapto ethanol). The bacterial lysate was centrifuged at 20000 g for 30 minutes to collect the soluble fraction. The soluble fraction was loaded on Ni-NTA column (Amersham). The column was washed twice with buffer containing 20 and 40 mM imidazol. The protein was eluted with 300 mM. The cleavage of N-terminal carrier polypeptide containing 6 histidine residues was performed as follows. After dialysis against thrombin cleavage buffer (20 mM Tris-HCl, 150 mM NaCl, 2.5 mM $CaCl_2$, pH 8.0) the cleavage was carried out for 24 h on room temperature using human thrombin (Sigma) at concentration of 5 U per mg of recombinant protein. The processed solution was loaded to Ni-NTA column and the EB1c protein was collected as the flow-through fraction. The homogeneity of the recombinant protein was checked by 15 % SDS-PAGE and the mass was confirmed by mass spectrometry.

Sample preparation and buffer. Pure protein samples were dialyzed against the standard working buffer (PBS: 10 mM sodium phosphate, 150 mM NaCl, pH 7.4). Protein concentration was determined by UV absorption at 280 nm in 6 M GdmCl [4]. The pH of samples solutions containing urea and GdmCl was adjusted after addition of the denaturant.

Circular dichroism (CD) spectroscopy. CD spectroscopy experiments were performed on Jasco J-715 and Jasco J-810 spectropolarimeters equipped with a computer-controlled water thermostat using quartz cuvettes of 0.1 and 1 cm optical path length. Spectra were recorded at scanning speed of 5 nm min^{-1} in 1 nm intervals. For isothermal denaturant-induced unfolding the samples were incubated overnight. The ellipticity at 222 nm (MRE_{222}) was sampled for three minutes. Thermal denaturation was performed by monitoring the change in MRE_{222} during continuous heating between 5 and 75-90 ℃ using heating rate of 1 ℃ min^{-1}. Reversibility was checked from the recovery of the CD signal after cooling, and was always at least 90%.

Analysis of spectroscopic data. The data obtained from thermal and isothermal chemical denaturation were analysed according to the 2-state model described by:

$$M + M \rightleftharpoons D \tag{3.1}$$

The equilibrium unfolding constant is defined as:

$$K_U = \frac{[M]^2}{[D]} \tag{3.2}$$

The fractions of monomer, f_M, and dimer, $f_D = 1 - f_M$, are defined as $\mathrm{f_M} = \frac{[\mathrm{M}]}{[\mathrm{M_0}]}$ and $\mathrm{f_D} = 1 - \mathrm{f_M} = \frac{2[\mathrm{D}]}{[\mathrm{M_0}]}$ respectively, where $\mathrm{M_0}$ is the known total protein concentration in monomer equivalents. Substitution into (3.2) yields:

$$K_U = \frac{[M]^2}{[D]} = \frac{2f_M^2 M_0}{1 - f_M} \tag{3.3}$$

The only physically meaningful solution of the quadratic equation 3.3 for $\mathrm{f_M}$ is:

$$f_M = \frac{-K_U + \sqrt{K_U^2 + 8K_U M_0}}{4M_0} \tag{3.4}$$

In analysis of quasi-sigmoidal denaturation curves it is usually assumed that the measured ellipticities in the pre-transitional and post-transitional segments correspond to the signals of the native state (dimer in case) and the denatured state (monomer in this case), respectively, and are linear functions:

$$\theta_D = \theta_{D,\gamma 0} + \alpha\gamma \tag{3.5}$$

$$\theta_M = \theta_{M,\gamma 0} + \beta\gamma \tag{3.6}$$

The terms γ indicate the independent variable, i.e. temperature or denaturant concentration, and γ_0 is the value at an arbitrarily chosen reference condition, usually 273.15 K and 0 M denaturant. The fraction monomer for each point of the experimental denaturation curve is given by:

$$f_M = \frac{\theta - \theta_D}{\theta_M - \theta_D} \tag{3.7}$$

In denaturation experiments the measured signal is assumed to represent a simple sum of the signals of the native and denatured states weighted in proportion to the progress of unfolding. Formally, this is algebralic rearrangement of equation 3.7:

$$\theta_i = f_M\theta_M + f_D\theta_D = \theta_D + f_M(\theta_M - \theta_D) \tag{3.8}$$

$\mathrm{f_M}$ is related to $\mathrm{K_U}$ and $\mathrm{M_0}$ according to equation 3.4. In thermal unfolding, the temperature dependence of K_U is given by the integrated van‘t Hoff equation:

$$K_U(T) = K_U(T_m)\ \exp\left\{\left[\frac{\Delta H_m}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right)\right]-\frac{\Delta C_p}{R\,T}\left[T-T_m-T\ \ln\left(\frac{T}{T_m}\right)\right]\right\} \tag{3.9}$$

Non-linear regression analysis to the combined equations 3.3-3.9 allows calculation of the unfolding enthalpy change, ΔH_m, and the unfolding heat capacity change, ΔC_p, at the denaturation midpoint, T_m. Since $f_M = 0.5$ at T_m, $K_U(T_m) = M_0$. In practice, ΔC_p is underdetermined in single denaturation curves. Statistically robust estimates for T_m and ΔH_m are obtained by eliminating the ΔC_p-containing term in equation 3.9. The approach is justified since ΔC_p is numerically small in comparison to ΔH_m (usually less than 5%). ΔC_p is calculated from the slope of ΔH_m-vs-T_m plots (Kirchoff‘s plots). An alternative way to obtain ΔH_m at T_m is to use the following equation [5]:

$$\Delta H_m = 6RT_m^2\left(\frac{\partial f_U}{\partial T}\right)_{T=T_m} \tag{3.10}$$

Data from urea and GdmCl isothermal unfolding experiments were analyzed using the linear-extrapolation-method (LEM; [6]). Equations 3.4-3.8 apply also in this type of experiments. In the framework of LEM the unfolding free energy at each denaturant concentration, $[D]$ is given by:

$$\Delta G_U(D) = -RTlnK_U(D) = \Delta \mathrm{G}_U^{H_2O} - m_{eq}[D] \tag{3.11}$$

$\Delta \mathrm{G}_U^{H_2O}$ is the unfolding free energy at 0 M denaturant and $m_{eq} = \frac{\partial \Delta G_U}{\partial [D]}$.

Differential scanning calorimetry (DSC). DSC experiments were carried out on a VP-DSC microcalorimeter (MicroCal). The calorimeter cell block is equipped with two cylindrical cells of 0.52 ml volume [7]. Pressure of 2 atm was applied on the adiabatic socket. Temperature scans from 5 to 90 ℃ were carried out at scanning rate of 1 ℃ min^{-1}. The protein was dialyzed overnight against PBS. The dialysis buffer was used to establish the instrumental baseline. After the subtraction of instrumental baseline, data were normalized for the molar concentration.

Analysis of DSC data. In principle, analysis of the temperature dependence of the moral heat capacity function is performed according to:

$$C_p(T) = C_{p,D} + f_M\Delta C_p + \Delta H_m\frac{df_M}{dT} = C_{p,D} + f_M\Delta C_p + \frac{K_U}{(1+K_U)^2}\frac{\Delta H_m^2}{RT^2} \tag{3.12}$$

$C_{p,D}$ is the heat capacity of the native state (dimer in the present case). f_M is defined as in equation 3.4, and the temperature dependence of K_U is expressed as in equation 3.9. As detailed in *Results and Descussion* section, the thermogram of EB1c is anomalous. Only the apparent excess heat capacity function, as defined by the last right-hand side term of equation 3.12, was subjected to analysis. It was constructed by connecting the ends of the heat absorption peak by a smooth spline function and subtracting this function from the thermogram. The calorimetric enthalpy, ΔH_{cal}, was calculated by numerical integration of the resulting excess heat absorption peak. The effective van't Hoff enthalpy was calculated by plotting the integrated area as function of the temperature and applying equation 3.10.

Protein stability curve. The functional form of the protein stability curve is given by the Gibbs-Helmholtz equation:

$$\Delta G_U(T) = \Delta H_m \left(\frac{1}{T} - \frac{1}{T_m}\right) + \Delta C_p \left(T - T_m - Tln\frac{T}{T_m}\right) - RTlnK_U(T_m) \tag{3.13}$$

If ΔH_m, T_m and ΔC_p are known from thermal melting experiments ΔG_U can be calculated at any temperature. Alternatively, data from isothermal chemical denaturation, i.e. $(\Delta G_U(T_i),T_i)$-pairs can be subjected to non-linear regression analysis according to equation 3.13 to optimize ΔH_m, T_m and ΔC_p. The best practice is to analyze the combined data from the two types of unfolding experiments.

Stopped flow folding/unfolding kinetics and data analysis Stopped-flow experiments were performed following time course of the MRE_{222} at 25 ℃ on a π*-180 instrument (Applied Photophysics). Path length was 10 mm and the slits width was set at 4 nm. Instrument dead time was 1-2 ms, and the mixing flow rate was 5 ml s^{-1}. Typically, five syringe firings were performed for each kinetic trace. Protein stock solutions were prepared by dissolving the protein either in PBS or in high concentration of denaturing agents (6 M GdmCl or 8 M urea). Refolding/unfolding was initiated by a rapid ten-fold dilution of the protein stock solution with solutions containing different amount of the denaturant to achieve the desired final protein concentration and denaturant concentration. Data analysis was done according to the 2-state model, as described in ref. [8]. The following set of equations was implemented for regression analysis:

$$b = \frac{2\sqrt{k_u^2}}{k_f\, M_0} \tag{3.14}$$

$$sq = \sqrt{b^2 + 8b} \tag{3.15}$$

$$z = \frac{4 + b - sq}{4 + b + sq}\, exp\{-0.5\, sq\, k_f\, t\, M_0\} \tag{3.16}$$

$$F = 0.25\, M_0 \left(\frac{-b + sq + b\, z + sq\, z}{1 - z}\right) \tag{3.17}$$

$$\theta(t) = \theta_0 + \theta_{max} \left(1 - \frac{F}{M_0}\right) \tag{3.18}$$

The microscopic rate constants for folding and unfolding at any denaturant concentration are $\mathrm{k_f}$ and $\mathrm{k_u}$, respectively. In the absence of kinetic complications (accumulation of intermediates, transition state movement) $\mathrm{k_f}$ and $\mathrm{k_u}$ depend linearly on [D] according to:

$$lnk_f = ln\mathrm{k}_f^{H_2O} + m_f[D] \tag{3.19}$$

$$lnk_u = ln\mathrm{k}_u^{H_2O} + m_u[D] \tag{3.20}$$

Parameters $\mathrm{m_f}$ and $\mathrm{m_u}$ (in units of $\mathrm{M^{-1}}$) describe the dependence of the logarithm of the rate constant on [D] in the general form m = dlnk/d[D]. Data transformation and analysis were performed using routines and in-house written scripts for the programs Origin (OriginLab), NLREG (Philip Sherod) and Mathematica (Wolfram Research).

3.3 RESULTS AND DISCUSSION

At concentrations higher than 20 μM (monomer equivalents) EB1c exists as a non-covalent dimer, as evidenced by gel filtration and multiangle light scattering (not shown). The far-UV CD spectrum has the typical spectral signature of a protein with high -helical content, in accord with structural data (not shown and ref. [2]).

3.3.1 Equilibrium unfolding

Spectroscopic analysis The result of a typical unfolding experiment using the temperature-induced changes in $\mathrm{MRE_{222}}$ as the experimental signal is shown in Fig. 3.2. With the usual assumption that the pre-transitional portion of the melting trace (5-40 ℃) corresponds to the native state (dimer), while the signal at T > 75 ℃ corresponds to the thermally-denatured state (monomer), and only these two states are populated in the transition zone, the data can be modeled with the combined eqs 3.1-3.8 (2-state model). The trace is described with $\mathrm{T_m} = 66.6 \pm 0.2$ ℃ and van't Hoff enthalpy change at that temperature, $\Delta\mathrm{H_m^{CD}} = 360 \pm 30$ kJ $\mathrm{mol^{-1}}$ ($\mathrm{C_{tot}} = 90$ μM monomer equivalents). Since $\Delta\mathrm{G_m^{CD}} = -\mathrm{RTlnK_U} = -\mathrm{RTlnC_{tot}} = 26.3$ kJ $\mathrm{mol^{-1}}$, the entropy change at $\mathrm{T_m}$ is $\Delta\mathrm{S_m^{CD}} = \frac{(\Delta\mathrm{H_m^{CD}} - \Delta\mathrm{G_m^{CD}})}{\mathrm{T_m}} = 950 \pm 100$ J $\mathrm{K^{-1}}$ $\mathrm{mol^{-1}}$.

Although it is theoretically possible to determine the unfolding heat capacity change, $\Delta\mathrm{C_p}$, from a single melting trace, the statistical significance of such estimates is low (due to the coupling within the parameter triplet ($\mathrm{T_m}$, $\Delta\mathrm{H_{vH}^{CD}}$, $\Delta\mathrm{C_p^{CD}}$)). With oligomeric proteins one can vary the protein concentration in order to introduce variation in $\mathrm{T_m}$ and $\Delta\mathrm{H_{vH}^{CD}}$, and calculate $\Delta\mathrm{C_p^{CD}} = \frac{\mathrm{d}\Delta\mathrm{H_{vH}^{CD}}}{\mathrm{dT_m}}$. However, the temperature variation is usually small to achieve reliable enthalpy variation. To overcome the problem, a series of melting experiments was performed in the presence of sub-denaturing amounts of urea (< 2.5 M urea; see below). The resulting Kirchoff's plot is shown in Fig. 3.3. The calculated heat capac-

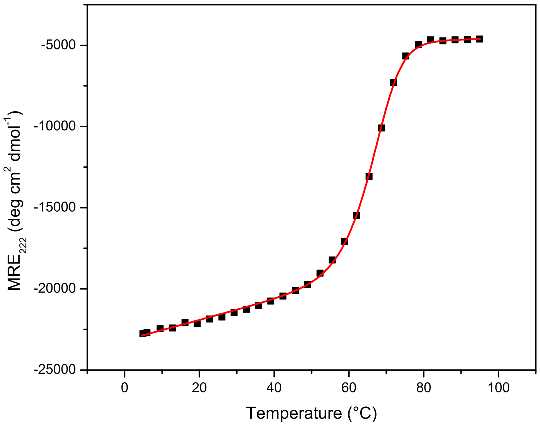


Fig. 3.2: Thermal melting of EB1c monitored by CD spectroscopy. Data with 90 μM monomer in standard buffer.The symbols are the measured MRE$_{222}$. The line is the best fit obtained according to eqs. 3.1 - 3.10. The calculated parameters are $T_m = 66.6 \pm 0.2$ ℃ and $\Delta H_m^{CD} = 360 \pm 10$ kJ mol^{-1}

ity change is $\Delta C_p^{CD} = 6.5 \pm 0.3$ kJ K^{-1} mol^{-1}. This number is surprisingly close to the expected ΔC_p for unfolding of an average protein of the size of EB1c (6.3 kJ K^{-1} mol^{-1}; [9]). Since all relevant parameters are known, the Gibbs-Helmhoz equation (eq 3.13)

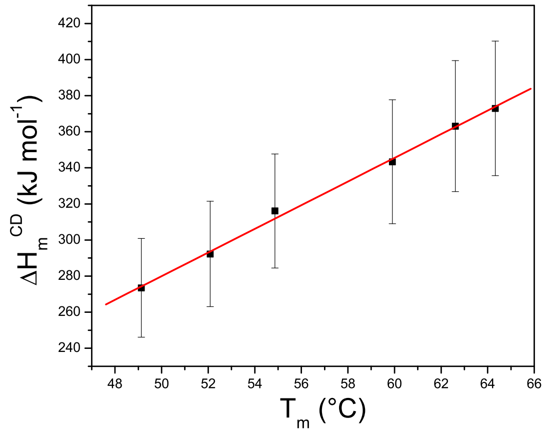


Fig. 3.3: Kirchoff's plot constructed from CD melting data. The points represent HmCD measured in the presence of sub-denaturing amounts of urea. The slope of the linear function describing the data is $\Delta C_p^{CD} = 6.5 \pm 0.3$ kJ K^1 mol^{-1}.

can be used to calculate ΔG_U at any temperature. At 25 ℃ $\Delta G_U = 50 \pm 5$ kJ mol^{-1}. The calculated stability curve is presented in Fig. 3.5. . The error in ΔG resulting

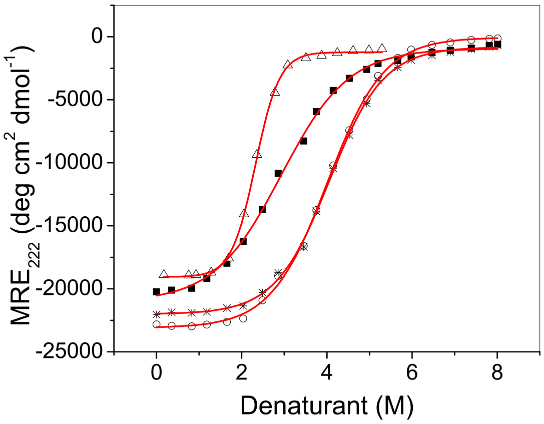


Fig. 3.4: Isothermal denaturant-induced unfolding of EB1c. Experiments were performed in standard buffer supplemented with increased amounts of urea at 5 ℃ (circles), 25 ℃ (asterisks) and 40 ℃ (squares) or of GdmCl at 25 ℃ (triangles). Protein concentration was 60 μM monomer (experiments in urea) and 40 μM monomer (experiments in GdmCl). The symbols are the measured MRE$_{222}$. The lines visualize the results of non-linear optimization analysis according to eqs 3.1-3.8

from the long extrapolation is large. Furthermore, the two-state unfolding model can not be taken for granted. For these reasons the stability of EB1c was assessed additionally by isothermal denaturant-induced unfolding experiments. Again, the experimental signal was $\mathrm{MRE_{222}}$. Unfolding curves were recorded at 5, 25 and 40 ℃ in urea and at 25 ℃ in GdmCl (Fig. 3.4). Regression analysis according to the combined eqs 3.1-3.8 assuming a 2-state model was performed to extract the values of the mid-points of denaturation, $\mathrm{D_{1/2}}$, the equilibrium m-value, $\mathrm{m_{eq}} = \frac{\mathrm{dG_U^D}}{\mathrm{d[D]}}$, and the unfolding free energy in plain buffer, $\Delta\mathrm{G_U^W}$. The data are summarized in Table 3.1 and the $\Delta\mathrm{G_U^W}$ values are shown in graphical form in Fig.3.5. As with $\Delta\mathrm{C_p}$, the $\mathrm{m_{eq}}$-values in both urea and GdmCl are within the range predicted by statistical analysis of the available database [9]. The error in ΔG resulting from the long extrapolation is large. Furthermore, the two-state unfolding model can not be taken for granted. For these reasons the stability of EB1c was assessed additionally by isothermal denaturant-induced unfolding experiments. Again, the experimental signal was $\mathrm{MRE_{222}}$. Unfolding curves were recorded at 5, 25 and 40 ℃ in urea and at 25 ℃ in GdmCl (Fig. 3.4).

The most robust way to construct the stability curve of a protein is to combine data from thermal unfolding (stability around $\mathrm{T_m}$) with data from isothermal unfolding at lower temperatures. To this end, (T, $\Delta\mathrm{G_U^W}$)-pairs from Table 3.1 were plotted together with the ($\mathrm{T_m}$, $\Delta\mathrm{G_U^{CD}}$)-pair obtained by thermal unfolding as described above ($\Delta\mathrm{G_U^W} = -\mathrm{RTlnK_U} = -\mathrm{RTlnC_{tot}} = 26.3\ \mathrm{kJ\ mol^{-1}}$ at 66.6 ℃). The data were subjected to regression analysis according to the Gibbs-Helmholtz equation (eq 3.13) to optimize simultaneously $\mathrm{T_m}$, $\Delta\mathrm{H_{vH}^{CD}}$ and $\Delta\mathrm{C_p^{CD}}$. The best three-parameter fit, which is illustrated in Fig. 3.5, was obtained with the following combination of values: $\mathrm{T_m^{fit}} = 66.2$ ℃ , $\Delta\mathrm{H_{vH}^{fit}} = 395 \pm 60\ \mathrm{kJ\ mol^{-1}}$ and $\Delta\mathrm{C_p^{fit}} = 6.7 \pm 2.2\ \mathrm{kJ\ mol^{-1}\ K_{-1}}$. At first glance, the agreement between the global fit parameters and the parameters from thermal unfolding is reasonable ($\frac{\mathrm{T_m^{CD}}}{\mathrm{T_m^{fit}}} = 1$, $\frac{\Delta\mathrm{H_m^{CD}}}{\Delta\mathrm{H_m^{fit}}} = 0.91$ and $\frac{\Delta\mathrm{C_p^{CD}}}{\Delta\mathrm{C_p^{fit}}} = 0.97$). However, as demonstrated in Fig. 3.5, the two sets of data systematically diverge below $\backsim$ 40 ℃, indicating either a systematic experimental error (which we are unable to identify) or (slightly) different unfolding mechanism, depending on the denaturing agent.

Calorimetric analysis. The thermal unfolding of EB1c was further characterized by DSC. Fig. 3.6 presents the thermogram recorded with 156 μM EB1c (monomer equivalents). Visual inspection of the trace reveals one main peculiarity: The heat capacity preceding the heat absorption peak increases very steeply with temperature. It is usually assumed that this portion of the trace represents the heat capacity of the native state. However, if this is valid for EB1c, the heat capacities of the native and denatured states cross at $\backsim$ 40 ℃ (lines a and b in Fig. 3.6A). It is known that such situation for a typically folded protein is very unlikely for physical reasons [10]. There are two possible explanations of the experimental observation of crossing baselines at low temperature.

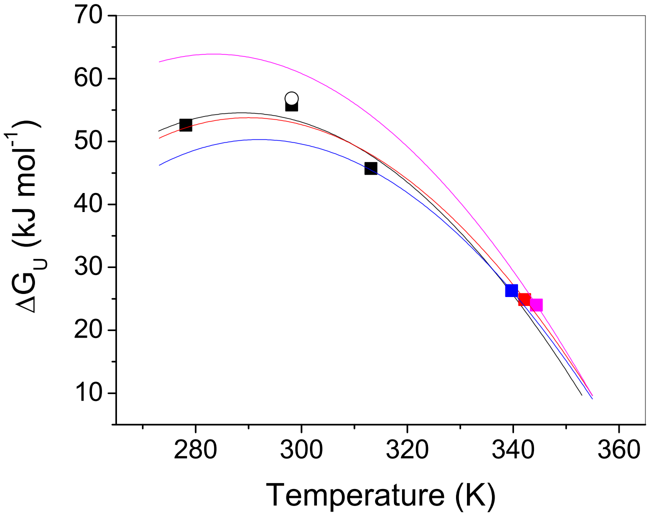


Fig. 3.5: Stability curves of EB1c. Differently colored ΔG_U-versus-T profiles (Gibbs-Helmholtz profiles according to eq 3.13) were constructed using different sets of data. Black line: Simultaneous fitting of data derived from isothermal urea-induced unfolding (black squares), GdmCl-induced unfolding (open circle), and thermal unfolding followed by CD (blue square). Best fit was obtained with $T_m^{fit} = 66.2 \pm 0.3$℃, $\Delta H_m^{fit} = 395 \pm 60$ kJ mol^{-1}, and $\Delta C_p^{fit} = 6.7 \pm 2.2$ kJ K^{-1} mol^{1}. Line and square in blue: Curve simulated with $T_m^{CD} = 66.6$℃ , $\Delta H_m^{CD} = 360$ kJ mol^{-1} , and $\Delta C_p^{CD} = 6.5$ kJ K^{-1} mol^{-1}, which are derived from thermal melting followed by CD. Line and square in red: Curve simulated with $T_{m,vH}^{DSC} = 69$℃ , $\Delta H_{m,vH}^{DSC} = 390$ kJ mol−1. Line and square in magenta: Curve simulated with $T_{m,cal}^{DSC} = 71$℃ , $\Delta H_{m,cal}^{DSC} = 460$ kJ mol^{-1}. The two latter curves were simulated with $\Delta C_p^{CD} = 6.5$ kJ K^{1} mol^{-1} . The blue, red and magenta squares, represent ΔG_U at 66.6 ℃ ($f_U = 0.5$ with 90 μM monomer; CD melting) , 69 ℃ ($f_U = 0.5$ with 156 μ M monomer; DSC melting) and 71 ℃ ($f_U = 0.59$ with 156 μ M monomer; at the temperature of the maximum heat absorption), respectively.

First, denaturation is incomplete, meaning that the heat capacity after completion of the main transition is lower than heat capacity of the significantly hydrated state lacking tertiary interactions. This can be excluded in the case of EB1c since the unfolding heat capacity (~ 6.5 kJ K^{-1} mol^{-1}) correlates well with the expected one based on the estimated increase of water accessible surface upon complete unfolding [11]. Second, the pre-transitional portion of the thermogram does not reflect the genuine heat capacity of the completely folded state. Rather, it contains contributions from temperature-dependent, gradual structural changes [12]. Indeed, line a in Fig. 3.6A is 0.347 kJ K^{-2} mol^{-1}, which translates into 18.8×10^{-3} kJ K^{-2} g^{-1}. This number is three times larger than the heat capacity of compact globular domains ($(6.7 \pm 1.7) \times 10^{-3}$ kJ K^{-2} g^{-1}; [9]). Line c in Fig. 3.6A represents the hypothetical molar heat capacity of EB1c, if it was an average protein. (The slope is 0.0067 10^{-3} kJ K^{-2} g^{-1} × 18412.2 g mol^{-1} = 0.123 kJK^{-2} mol^{-1}. We have arbitrarily equaled the calculated function with the experimental heat capacity at ~ 10 ℃ , assuming that the lower the temperature, the more compact the protein is. Even if this is an ungrounded speculation, it helps to understand the shape of the thermogram. The hypothetical unfolding heat capacity change, i.e. the difference ΔC_p = (line b) - (line c) varies between 7 kJ K^{-1} mol^{-1} (at 10 ℃) and 5 kJ K^{-1} mol^{-1} (at 80 ℃), thus being close

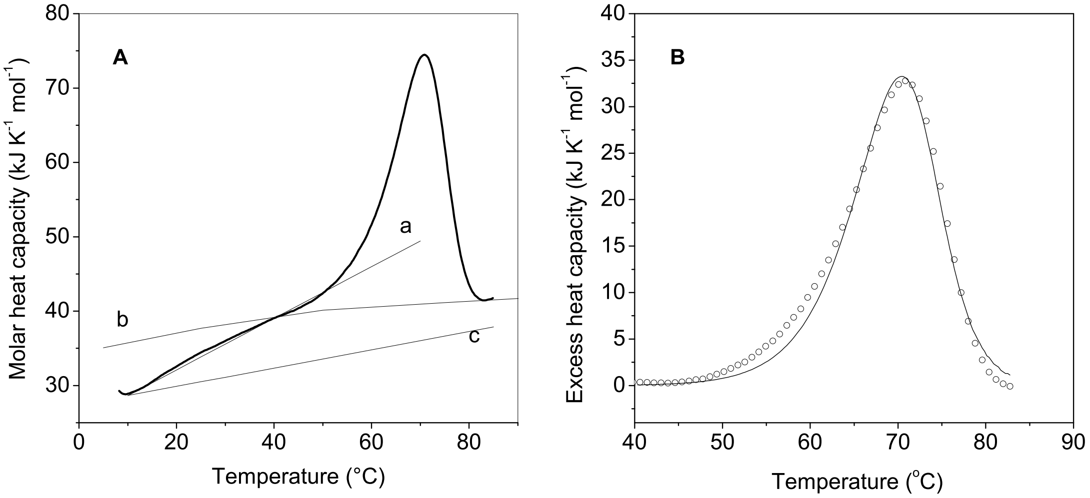


Fig. 3.6: Thermal melting of EB1c monitored by DSC. (Panel A) Thermogram recorded with 156 μM EB1c (monomer equivalents) in standard buffer (thick solid line). The thin lines a-c represent the apparent heat capacity of the native state (a), the heat capacity of the denatured state calculated from the amino acid sequence (b), and the theoretical heat capacity expected for an average native protein of the size of EB1c. For details, see the main text. Due to the relatively low solubility of EB1c precluded measurement of the absolute heat capacity. The trace was shifted on the Y-axis to match line b at 80 ℃. (Panel B) Excess heat capacity function. The symbols represent the experimental data after subtraction of a smooth spline function connecting the heat capacity trace between 40 and 80 ℃. The solid line is the best fit to the data according to a model of unfolding associated with subunit dissociation (last term of eq 3.12).

to the non-calorimetric estimate of 6.5 ± 2 kJ K^{-1} mol^{-1} (see above). It follows that the gradual pre-transition heat absorption likely mirrors either partial non-cooperative unfolding, or rapid inter-conversion of intermediary states accompanied by intense enthalpy fluctuations, or very intensive thermal fluctuations.

The excess heat capacity function was analyzed to extract the enthalpy of unfolding. The area of the peak shown in Fig. 3.6B represents the model independent calorimetric estimate, $\Delta\mathrm{H}^{\mathrm{DSC}}_{\mathrm{m,cal}} = 460$ kJ mol^{-1}. From the shape of the peak, the effective van't Hoff enthalpy is $\Delta\mathrm{H}^{\mathrm{DSC}}_{\mathrm{m,vH}} = 393$ kJ mol^{-1} (see eq 3.10). (We do not have enough data to estimate the errors of the given numbers, but assume that the principal error comes from the uncertainty in protein concentration determination, which is ~ 5 % in this case). The ratio $\dfrac{\Delta\mathrm{H}^{\mathrm{DSC}}_{\mathrm{m,vH}}}{\Delta\mathrm{H}^{\mathrm{DSC}}_{\mathrm{m,cal}}}$ is thus 0.85 and indicates that the unfolding process deviates from the 2-state model, albeit not by much. This conclusion is further supported by modeling of the excess heat capacity according to 2-state equilibrium between folded dimer and unfolded monomer (eq 3.12). As seen in panel B of Fig. 3.6, the model captures the main characteristics of the process (the typical asymmetry of oligomeric transitions) but does not completely explain the experimental data, especially the heat capacity at the beginning of the unfolding transition.

3.3.2 Folding/unfolding kinetics

The rates of folding and unfolding of EB1c were studied by following the time course of helix formation/disruption after rapid dilution from/into denaturant at 25 °C. The folding experiments were performed in the range 0.2-1.9 M GdmCl and 0.4-3.4 M urea. Unfolding was characterized between 2.7 and 5.2 M GdmCl, and between 3.5 and 7.6 M urea. Representative kinetic traces are shown in Fig. 3.7. Folding and unfolding are both described precisely by single kinetic phases. All data can be fit with equations assuming the presence of only folded diner and unfolded monomer (eqs 3.14 - 3.18 ; 2-state model). As shown in Fig. 3.8, in both denaturants the folding and unfolding limbs of the Chevron plot do not reveal any statistically significant curvature. Equations 3.19 and 3.20 were fit to the data to calculate the kinetic constants extrapolated to zero denaturant and the

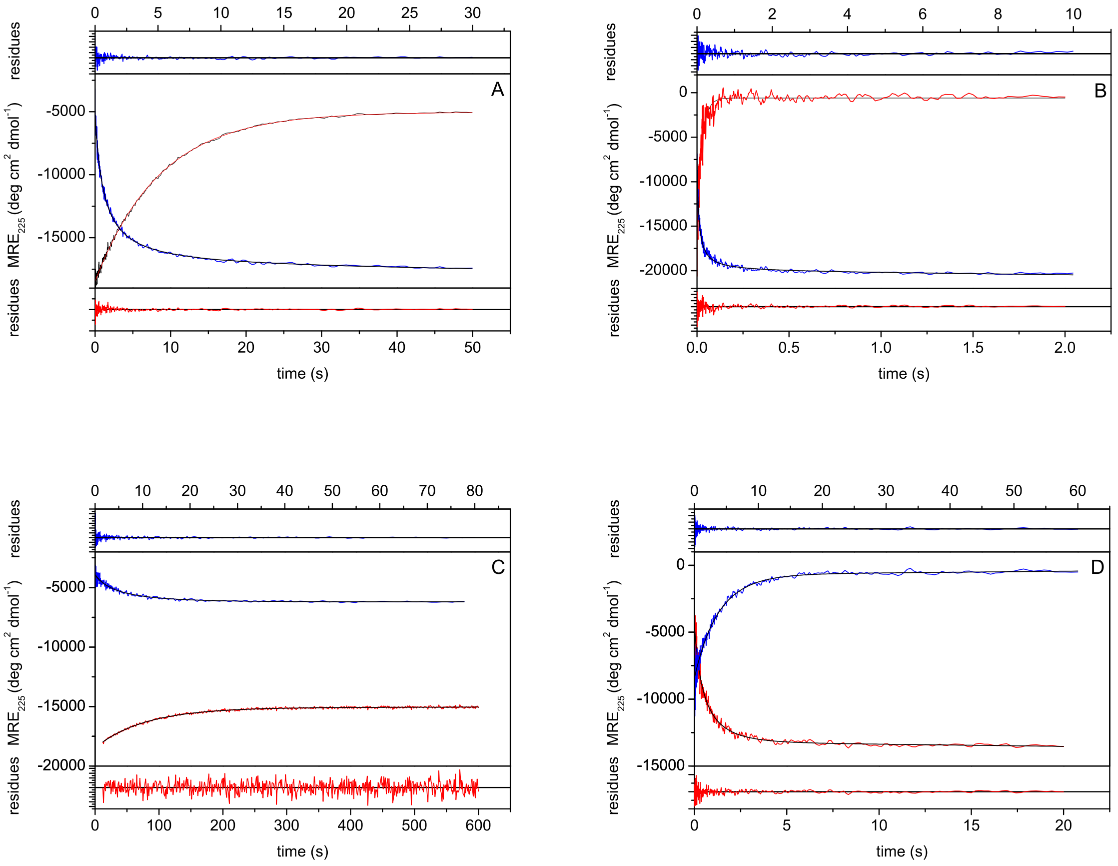


Fig. 3.7: Representative kinetic traces for folding and unfolding of EB1c. As an illustration, the fastest and the slowest reactions measured in GdmCl (panels A and B) and in urea (panels C and D) are shown. In all panels, folding traces are shown in blue, and unfolding traces are shown in red. (Panel A) Folding in 1.59 M GdmCl (15.4 μM EB1c) and unfolding in 2.8 GdmCl. (Panel B) Folding in 0.42 M GdmCl (15.4 μM EB1c) and unfolding in 5.4 M GdmCl. (Panel C) Folding in 3.25 M urea (20.2 μM EB1c) and unfolding in 3.8 M urea. (Panel D) Folding in 0.7 M urea (20.2 μM EB1c) and unfolding in 7.5 M urea. The residuals in the optimization procedure to extract the values of k_f and k_u (eqs 3.14-3.18) are shown above and below each panel and are colored correspondingly.

kinetic m-values, $m_f = \dfrac{\mathrm{dln}k_f}{\mathrm{d[D]}}$ and $m_f = \dfrac{\mathrm{dln}k_u}{\mathrm{d[D]}}$. The kinetic parameters are summarized in Table 3.1 .

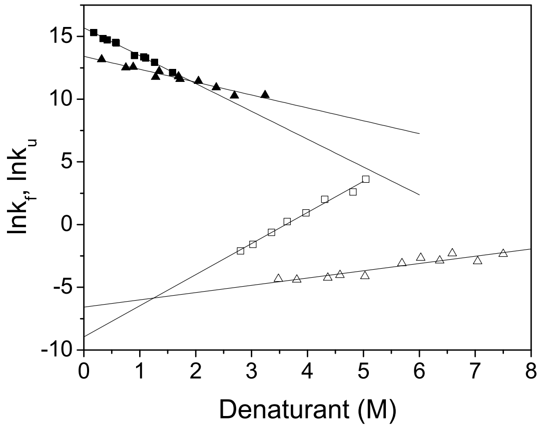


Fig. 3.8: Denaturant dependence of the rates of folding (filled symbols) and unfolding (open symbols) of EB1c. Data in GdmCl are represented with squares; data in urea are presented with triangles. The experiments were performed at 25 ℃ in standard buffer supplemented with denaturant. Final concentrations were 11.5-21.2 μM (monomer equivalents). The lines are best fits according to eqs 3.19 and 3.20 and are listed in Table 3.1 .

Tab. 3.1: Data from isothermal denaturant-induced unfolding of EB1c and kinetic unfolding/refolding experiments.[a]

Temperature (℃)	$D_{1/2}$ (M)	m_{eq} (kJ mol^{-1} M^{-1})	ΔG_U^D (kJ mol^{-1})
5[b]	3.9	6.4	52.6
25[b]	3.9	7.3	55.8
25[c]	2.3	12.7	56.8
40[b]	2.6	6	45.7

Denatur. (25 ℃)	k_f (M^{-1} s^{-1})	k_u (s^{-1})	$\Delta G_U^{\mathrm{kin}^d}$ (kJ mol^{-1})	m_f (M^{-1})	m_u (M^{-1})	m_{kin}^e (kJ mol^{-1} M^{-1})	β_T^f
urea	6.0×10^5	25×10^{-3}	47.8	-1.02	0.56	3.9	0.65
GdmCl	66.0×10^5	1.4×10^{-3}	60.9	-2.20	2.50	11.7	0.47

[a] All experiments in standard buffer at the indicated temperature. Protein concentration was 60 μM monomer (experiments in urea) and 40 μM (experiments in GdmCl). The estimated errors are in the range of ± 0.1 M ($D_{\frac{1}{2}}$), ± 0.3 kJ mol^{-1} M^{-1} (m_{eq}), and ± 3-5 kJ mol^{-1} (ΔG_U^D).

[b] Experiments in urea.

[c] Experiment in GdmCl.

[d] $\Delta G_U^{\mathrm{kin}^d} = -RT\ln\frac{k_u}{k_f}$

[e] $m_{kin} = RT(|m_{on}| + |m_{off}|)$

[f] $\beta_T = 1 - \frac{m_u}{|m_f|+|m_u|}$

3.3.3 The available data point to a complicated folding/unfolding mechanism

Folding/unfolding of many small protein domains proceeds without detectable intermediates, the relative population of folded and unfolded states depending on the solvent conditions. This 2-state model does not imply that there are no intermediates at all. It represents an approximation stating that the molecular species (conformations) can be divided into two groups, which are separated by a high-energy barrier: mainly disordered monomers and mainly ordered (folded) dimers. Energetic differences within the states (caused by conformational fluctuations) are considered as small in comparison to the energy difference/barrier between the states. Although evidence has been found that intermediates can be populated in folding of some coiled coils [13, 14], most of the available data suggest a simple folding/unfolding mechanism [15–18]. Experimentally, the validity of the 2-state approximation is examined according to the following criteria. *(i)* The unfolding free energy and the folding/unfolding kinetics are identical within error when assessed by different experimental signals, e.g. fluorescence, far-UV, near-UV CD spectroscopy, etc. *(ii)* The effective, model-dependent unfolding enthalpy describing a 2-state conformational transition coincides with the calorimetric, model-independent parameter. *(iii)* The equilibrium unfolding constant measured in an equilibrium unfolding experiment equals the unfolding constant calculated as $\mathrm{K_{U,kin}} = \frac{\mathrm{k_u}}{\mathrm{k_f}}$, $\mathrm{k_u}$ and $\mathrm{k_f}$ being the microscopic rate constants for unfolding and folding, respectively. The equality between the equilibrium and kinetic m-values must hold, i.e. $\mathrm{m_{eq}} = \mathrm{m_{kin}} = \mathrm{RT}(\mid \mathrm{m_f} \mid + \mid \mathrm{m_u} \mid)$. We cannot evaluate the validity of *(i)*, since in the present work only CD experiments were performed. In the following we examine the validity of *(ii)* and *(iii)*.

As already mentioned, the ratio $\frac{\Delta \mathrm{H^{DSC}_{m,vH}}}{\Delta \mathrm{H^{DSC}_{m,cal}}} = 0.85$ points to the presence of some intermediate states in the transition region. It should be noted, that $\Delta \mathrm{H^{DSC}_{m,cal}}$, was calculated using a molecular mass of 18.412 kDa, corresponding to the full length construct. However, the C-terminal 15-17 residues are not seen in the crystal structure and are likely disordered. $\Delta \mathrm{H^{DSC}_{m,cal}}$ recalculated with the molecular mass of the portion of the protein which is crystallographically well defined (14.314 kDa) is 590 kJ mol^{-1}, and consequently $\frac{\Delta \mathrm{H^{DSC}_{m,vH}}}{\Delta \mathrm{H^{DSC}_{m,cal}}}$ becomes smaller than 0.7, a quite significant deviation from the 2-state model. With the recalculated unfolding enthalpy, the discrepancy between the stability curves obtained from calorimetric data and from denaturant induced unfolding will become even more pronounced (see Fig. 3.5). Altogether, the calorimetric data suggest that the folded state is subjected to some structural changes (or otherwise intense thermal fluctuations) that start at rather low temperatures and exhibit low cooperativity. These changes continue also in the transition zone and overlap with the main unfolding event, which is accompanied by subunit dissociation. Most likely, the N-terminal coiled coil sub-domain is involved in gradual loosening of packing interactions. It has been demonstrated that the EB1c coiled coil tail is only half folded at ∼ 10 ℃ and undergoes non-cooperative thermal unfolding [2].

The equilibrium and kinetic parameters of EB1c denaturant-induced folding/unfolding at 25 ℃ are compared in Table 3.1 and in graphical form in Fig. 3.9. Both, ΔG_U and m-values, extracted either at equilibrium conditions or from kinetic data, are in reasonable agreement if the denaturant was GdmCl. The difference of 4 kJ mol^{-1} in $\Delta G_U (< 5\ \%)$ and 1 kJ mol^{-1} M^{-1} in m-values (∼ 8 %) can be considered to be within the experimental error.

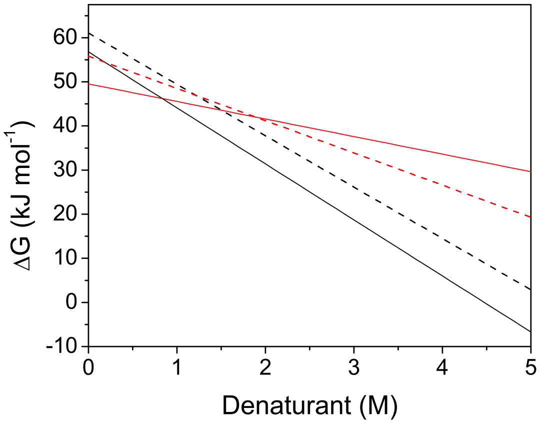


Fig. 3.9: Denaturant dependence of the unfolding free energy of EB1c. The lines representing the $\Delta G_U^D = f([D])$ functions derived from equilibrium data (solid) and from kinetic data (dashed) were calculated with the data from Table 3.1 . Black, data in GdmCl; red, data in urea.

In sharp contrast, the data in urea deviate by ∼15 % and ∼50 %(ΔG_U) and m-values, respectively. There is one important difference between the data sets collected in GdmCl and urea. As illustrated in Fig. 3.10, while the kinetic amplitudes (ΔMRE_{225}) for folding and unfolding are identical (± 5 %) to the amplitudes observed for equilibrium unfolding in GdmCl, the amplitudes for folding largely underestimate the equilibrium amplitudes in urea. (The amplitudes in the unfolding branch of the urea-derived Chevron plot match those measured at equilibrium conditions, in accord with the common knowledge that unfolding at high denaturant concentrations is a simple process [19]). Fig.3.7 shows that in the typical time window of a stopped-flow refolding experiment (10-80 s in this case) the kinetic traces in urea can be precisely described by a bimolecular association kinetic model. It follows that refolding in urea passes through a kinetic trap, the escape from which requires considerable time. In one scenario, the folding pathway is branched: Part of the unfolded ensemble (its population being dependent on the denaturant concentration) rapidly reaches the native state, while another part transverses a more complicated, slow pathway. In an alternative scenario, there is a single folding pathway, in which the total population escapes the kinetic trap by monomolecular and time-consuming rearrangement(s). Indeed, the second mechanism appears more realistic. It should be noted that the complications are seen at intermediary urea concentrations and are possibly caused by the combination of the rates of folding and unfolding. The

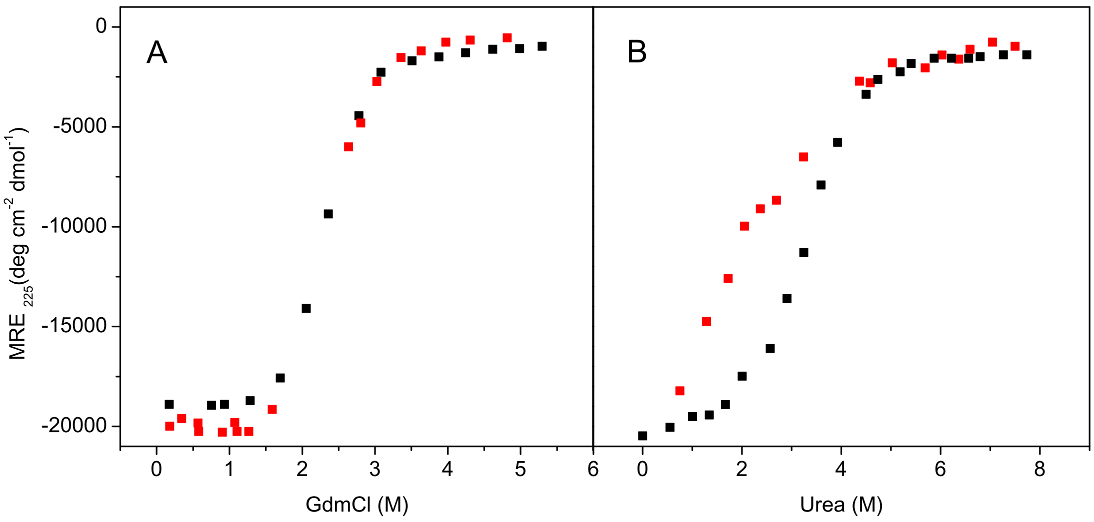


Fig. 3.10: Equilibrium and kinetic amplitudes observed in denaturant-induced experiments at 25 °C. (Panel A) Data in GdmCl. (Panel B) Data in urea. In both panels equilibrium amplitudes are shown in black, and kinetic amplitudes are shown in red.

observation has no implications for the folding mechanism at benign conditions. In such a view, the rate constant, k_f, measured in refolding experiments in urea is an apparent bimolecular constant describing formation of the intermediate. Due to baseline instability it was impossible to detect and quantify the slow kinetic phase. It should be noted that far-UV CD experiments at 225 nm report only on the denaturant-dependent degree of formation/disruption of secondary structure elements (α-helices for an all-α protein like EB1c). Therefore, any interpretations of the kinetic folding pathway of EB1c are devoid of insights into processes involving formation of tertiary contacts. Nevertheless, as far as the m-value correlates with the denaturant-sensitive surface exposed upon unfolding, a m_{kin} value much smaller than m_{eq} possibly indicates that the proposed intermediate buries much less molecular surface then the native dimer. Further experiments are required to resolve the problem.

The very different kinetic behavior of EB1c in GdmCl and urea must be the consequence of the chemical nature of the denaturants. Discrepant results on protein stability determination in the two types of experiments have been reported [20, 21]. As discussed in more details in Chapter 2, GdmCl acts as a dissociated salt and screens charge-charge interactions. We suggest that in the case of EB1c electrostatic effects are dominating. The protein contains 14 aspartic acids, 30 glutamic acids, and only 6 lysines and 6 arginines. This means that at neutral pH the protein is heavily negatively charged (theoretical pI is 3.9). Electrostatic repulsion is expected to slow down the folding rate in different ways: by decelerating the diffusion of the negatively charged polypeptide chains towards each other, by opposing the folding/consolidation within the initial encounter complex, and possibly by destabilizing the transition state. In the presence of GdmCl these effects are likely attenuated. As shown in Fig. 3.8 the folding below 2 M denaturant is faster in GdmCl than in urea. Screening of opposite charges might facilitate faster escape from the putative kinetic trap proposed above. Inspection of Fig. 3.8 also shows that at low denaturant concentration unfolding in GdmCl is slower than in urea. Clearly, the extrapolation from the transition region is a long one, but the observation makes sense, since screening of repulsive charges will stabilize the folded state. Work is under way in this laboratory to investigate the role of charge-charge interactions to the stability and folding of EB1c. Preliminary results at pH 2, where most of the negatively charges are neutralized, are in line with the discussed proposition. The midpoint of urea denaturation is higher by 3 M at pH 2 and $\Delta\Delta G_U = \Delta G_U^{pH2} - \Delta G_U^{pH7} \approx 15$ kJ mol^{-1} (not shown).

In conclusion, the presented data force the interpretation that folding/unfolding of EB1c is not a simple conformational transition between completely unfolded monomers and completely folded dimers. The premature stage of analysis notwithstanding, the likely reason is the uncoupling of formation/disruption of the N-terminal coiled coil and the C-terminal four-helix bundle. Tentatively, we attribute the uncoupling to electrostatic effects. Visual analysis of the EB1c X-ray structure reveals two spots where strong electrostatic repulsion between negatively charged groups is expected: *(i)* the N-terminus of the coiled coil sub-domain (shortest $O^\gamma - O^\gamma$ distances being 3.6 to 5.4 Å) and *(ii)* the loop region connecting the $\alpha 1$ and $\alpha 2$ helices (shortest $O^\gamma - O^\gamma$ distances being 2.7 to 5.4 Å). Since the coiled coil sub-domain is unstable in isolation [2], the four-helix bundle could remain folded at elevated temperatures and could fold rapidly/unfold slowly at medium urea concentrations, meaning that any intermediary states grossly represent a folded four-helix bundle and unfolded coiled coil. However, the extreme local concentration of negative charges in the loop region invokes in mind another scenario: The coiled coil forms rapidly and is kinetically quasi-stable in a broad temperature range and denaturant concentration range, while the main energetic barrier separating the unfolded (or any intermediary species) and the native state is crossed upon chain-orientation reversal, which is highly disfavored by charge-charge repulsion. One should note that the local concentration of repulsive charge-charge interactions around the loop region is probably higher due to the supposed, entropy-driven collapse of the heavily negatively charged unfolded stretches adjacent to the end of the 2 helices. In any case, shielding of repulsive charge-charge interactions is expected to decrease both the stability and the kinetic differences between the two sub-domains, just as the experiment evidences.

3.3.4 Stability of EB protein isoforms

Three isoforms of the EB protein have been identified so far [1]. The exact function of EB2 and EB3 isoforms is not yet completely understood. Interestingly, however, *in vivo* experiments demonstrated that heterodimerization takes place between EB1 and EB3, while EB2 is inert in this respect. Native gel electrophoresis experiments show that this behavior can be reproduced with the C-terminal domains (EBXc) *in vitro* (Christian de Groot and Michel Steinmetz, personal communication). The sequence of the three isoforms is quite conserved, especially in pair-wise comparisons (Fig. 3.11). Interestingly,

hEB1c:
DE-AAEL**M**QQ-V**NV**LKLT-VE**D**LEKE-RDFYF**G**K-LRNIELI-CQENE**G**EN**DP**VLQ**R**IVDILYATDE G**F**VI PD E**GGPQ**   EEQ   EEY

hEB2c:
TQ-VIQL**N**EQ-V**HS**LKLA-LE**G**VEKE-RDFYF**G**K-LREIELL-CQEHG**Q**EN**DD**LVQ**R**LMDILYASEEHEG**H**TEEPEAE**EQAH**   EQQPPQQEEY

hEB3c:
AQ-*I*LEL**N**QQ-L**VD**LKLT-VD**G**LEKE-RDFYF**S**K-LRDIELI-CQEHE**S**EN**SP**VIS**G**IIGILYATEE G**F**A PP E**DDEI**E**EH**QQEDQ   DEY

Fig. 3.11: Sequence of EB protein isoforms. Positions with non-conservative substitutions are in bold. Note that in almost all cases only one of the three isoforms contains a substitution. The sequence of the folded protein (according to X-ray data) is underlined.

the different length of the C-terminal unstructured tail comprises the main difference between the isoforms. Simple sequence and structural analyses reveal no clues about the inability of EB2c to participate in heterodimerization reactions. As a first step toward a comparative biophysical characterization of the three proteins thermal melting curves were recorded by CD spectroscopy (Fig. 3.12A). Clearly, the proteins possess different

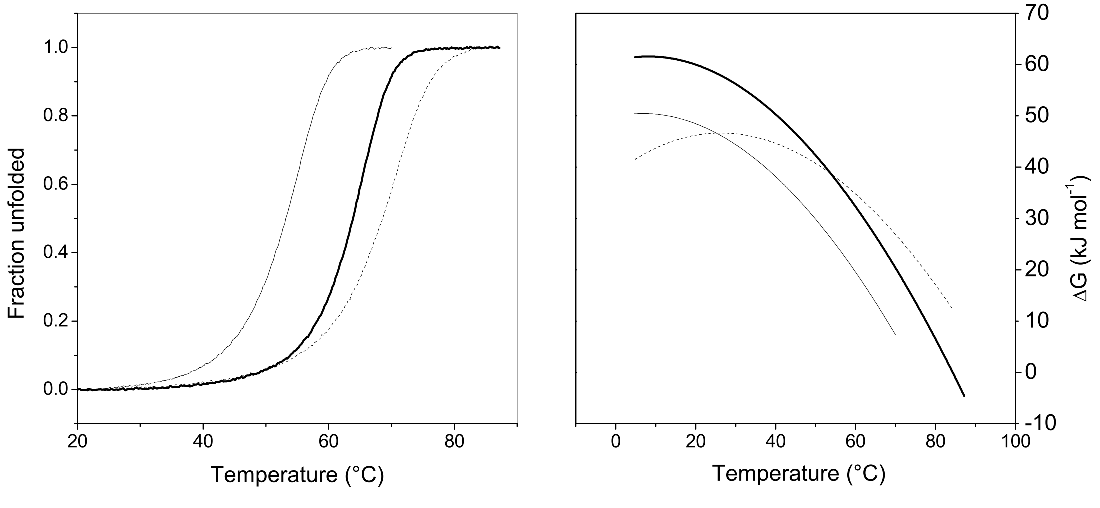


Fig. 3.12: Stability of EB protein isoforms. (Panel A) Thermal melting curves of EB1c (thick line), EB2c (dashed line) and EB3c (thin line) were recorded by CD with 50 μM protein (monomer equivalents) in standard buffer. (Panel B) Stability curves calculated according the eq 3.13, assuming that ΔC_p does not differ for the 3 isoforms.

thermal stability. The measured unfolding enthalpies are 400 kJ mol^{-1} at 63.8 ℃ (EB1c), 320 kJ mol^{-1} at 68.5 ℃ (EB2c), and 350 KJ mol^{-1} at 53.1 ℃. Very interestingly, EB2c unfolds with the lowest enthalpy and the highest melting temperature, indicating severely reduced cooperativity of the transition. To compare the thermodynamic stabilities in the relevant temperature range, the stability curves were simulated according to eq 3.13, making the reasonable assumption that the unfolding heat capacity increments are compatible. Fig. 3.12B demonstrates that EB1c is the most stable isoform below 50 ℃ Due to the low unfolding enthalpy, EB2c is less stable, and is isostable with at $\sim$ 25 ℃, in spite of the 15 ℃ difference in their melting temperatures. Work is progress to complete the stability profiles with data from chemical unfolding and to determine the rates of folding and unfolding. Differences in the biophysical properties of the three isoforms might be directly related to their biological function.

BIBLIOGRAPHY

[1] Carvalho, P, Tirnauer, J. S, & Pellman, D. (2003) Surfing on microtubule ends. *Trends Cell Biol.* **13**, 229–237.

[2] Honnappa, S, John, C. M, Kostrewa, D, Winkler, F. K, & Steinmetz, M. O. (2005) Structural insights into the EB1-APC interaction. *EMBO J.* **24**, 261–269.

[3] Hayashi, I & Ikura, M. (2003) Crystal structure of the amino-terminal microtubule-binding domain of end-binding protein 1 (EB1). *J. Biol. Chem.* **278**, 36430–36434.

[4] Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948–1954.

[5] Marky, L. A & Breslauer, K. J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **26**, 1601–1620.

[6] Schellman, J. A. (2002) Fifty years of solvent denaturation. *Biophys. Chem.* **96**, 91–101.

[7] Plotnikov, V. V, Brandts, J. M, Lin, L. N, & Brandts, J. F. (1997) A new ultrasensitive scanning calorimeter. *Anal. Biochem.* **250**, 237–244.

[8] Milla, M. E & Sauer, R. T. (1994) P22 Arc repressor: folding kinetics of a single-domain, dimeric protein. *Biochemistry* **33**, 1125–1133.

[9] Myers, J. K, Pace, C. N, & Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein. Sci..* **4**, 2138–2148.

[10] Naganathan, A. N, Perez-Jimenez, R, Sanchez-Ruiz, J. M, & Muoz, V. (2005) Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry* **44**, 7435–7449.

[11] D'Aquino, J. A, Gmez, J, Hilser, V. J, Lee, K. H, Amzel, L. M, & Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25**, 143–156.

[12] Carra, J. H & Privalov, P. L. (1996) Thermodynamics of denaturation of staphylococcal nuclease mutants: an intermediate state in protein folding. *FASEB J.* **10**, 67–74.

[13] Balakrishnan, G, Hu, Y, Case, M. A, & Spiro, T. G. (2006) Microsecond melting of a folding intermediate in a coiled-coil peptide, monitored by T-jump/UV Raman spectroscopy. *J. Phys. Chem. B.* **110**, 19877–19883.

[14] Dragan, A. I & Privalov, P. L. (2002) Unfolding of a leucine zipper is not a simple two-state transition. *J. Mol. Biol.* **321**, 891–908.

[15] Dürr, E & Jelesarov, I. (2000) Thermodynamic analysis of cavity creating mutations in an engineered leucine zipper and energetics of glycerol-induced coiled coil stabilization. *Biochemistry* **39**, 4472–4482.

[16] Bosshard, H. R, Dürr, E, Hitz, T, & Jelesarov, I. (2001) Energetics of coiled coil folding: the nature of the transition states. *Biochemistry* **40**, 3544–3552.

[17] Litowski, J. R & Hodges, R. S. (2002) Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity. *J. Biol. Chem.* **277**, 37272–37279.

[18] Moran, L. B, Schneider, J. P, Kentsis, A, Reddy, G. A, & Sosnick, T. R. (1999) Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking. *Proc. Natl. Acad. Sci. U. S .A.* **96**, 10699–10704.

[19] Fersht, A. (1998) Structure and Mechanism in Protein. Sci.ence: A Guide to Enzyme Catalysis and Protein Folding. (Freeman, W. H., Publishers).

[20] Akhtar, M. S, Ahmad, A, & Bhakuni, V. (2002) Guanidinium chloride- and urea-induced unfolding of the dimeric enzyme glucose oxidase. *Biochemistry* **41**, 3819–3827.

[21] Bjelić, S, Karshikoff, A, & Jelesarov, I. (2006) Stability and folding/unfolding kinetics of the homotrimeric coiled coil Lpp-56. *Biochemistry* **45**, 8931–8939.

CHAPTER 4

PROBING THE ENERGETIC AND KINETIC IMPACT OF TOPOLOGICALLY CONSERVED CONTACTS IN THE SIV GP41 SIX-HELIX BUNDLE

4.1 INTRODUCTION

The surface proteins of immunodeficiency viruses (HIV in humans, SIV in simian) play a key role in the early state of virus infection [1, 2]. They are involved in docking of the virus to the cell surface. The glycoprotein 41 (gp41) is synthesized as a gp160 precursor protein, which is subsequently cleaved to produce two noncovalently associated subunits, gp120 and gp41. The surface subunit gp120 determines viral tropism through interaction with the primary cellular receptor CD4 and particular chemokine receptors. The transmembrane subunit gp41 mediates direct fusion of the viral envelope with the cellular membrane. By analogy with the structural changes in the influenza virus hemagglutinin (HA) protein, it has been postulated that a large-scale structural rearrangement of gp41 is involved in the transition of the metastable native (prefusogenic) state to a stable fusion-active (fusogenic) state [2, 3].

The extracellular ectodomain of gp41 exhibits domain organization with several characteristic features which likely determine its role during activation of retroviral membrane fusion (see Fig. 4.1A). At the N-terminus there is a hydrophobic, glycine-rich sequence, referred to as the fusion peptide (fp), which is thought to insert into the cellular membrane at an early step of the fusion process [4]. The protein contains two regions with a 4,3 hydrophobic heptad repeat predicted to form coiled-coil structures [2]. The N-terminal heptad-repeat region is immediately C-terminal to the fusion peptide, while the C-terminal heptad-repeat region is located adjacent to the viral membrane (see Fig. 4.1A). The twoheptad-repeat regions are seqentially connected by a loop region containing a cystine link. Structural studies demonstrated that peptides corresponding to the N-terminal and C-terminal heptad repeats (N and C peptides), either non-linked or connected by a short linker, associate to form a stable, α-helical trimer of antiparallel heterodimers

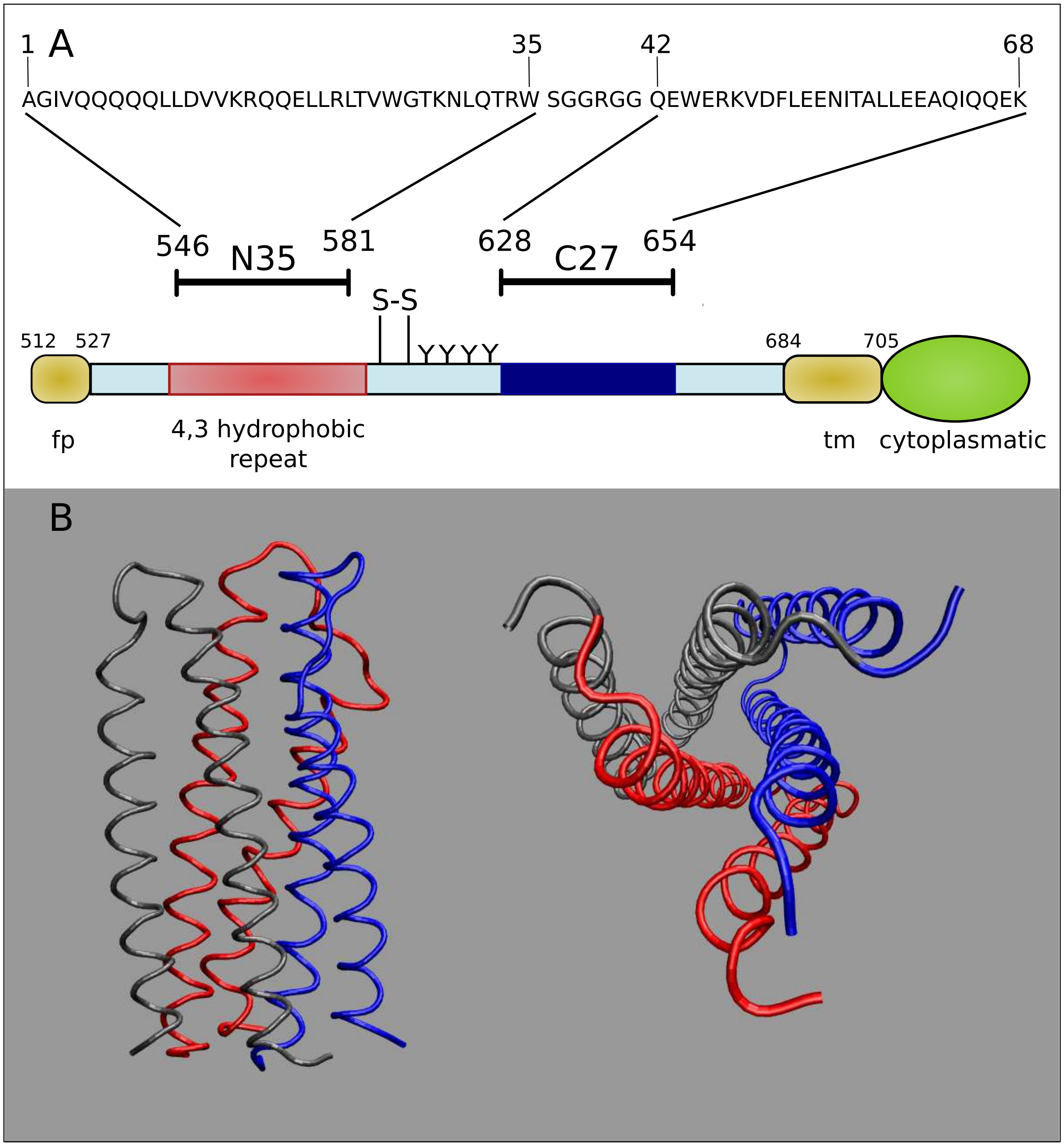


Fig. 4.1: The gp41 ectodomain core. (Panel A) A schematic representation of gp41. The important features of the gp41 ectodomain and the amino acid sequences of the N35 and C27 peptides are shown. N35(L6)C27 consists of N35 and C27 plus a linker of six hydrophilic residues. The disulfide bond and four potential N-glycosylation sites are depicted. The residues are numbered according to their position in SIV_{mac} gp160. (Panel B) The N35(L6)C27 crystal structure of the HIV-1 gp41 ectodomain core. The left panel shows a side view of the N35(L6)C27 trimer. The right panel shows an end-on view looking down the 3-fold axis of the trimer toward loops. The C27 helices pack against the surface of the internal trimeric N35 coiled-coil.

[2, 5, 6]. Three N helices form a central three-stranded coiled-coil, whereas three C helices pack in the reverse direction into highly conserved hydrophobic grooves at the surface of the N-terminal coiled coil ([2]; see Fig. 4.1B). It has been proposed that this structure likely represents the fusion-active state of gp41 [2, 3, 5, 6]. The detailed mechanism of gp41-mediated membrane fusion is still unknown. Nevertheless, the available information

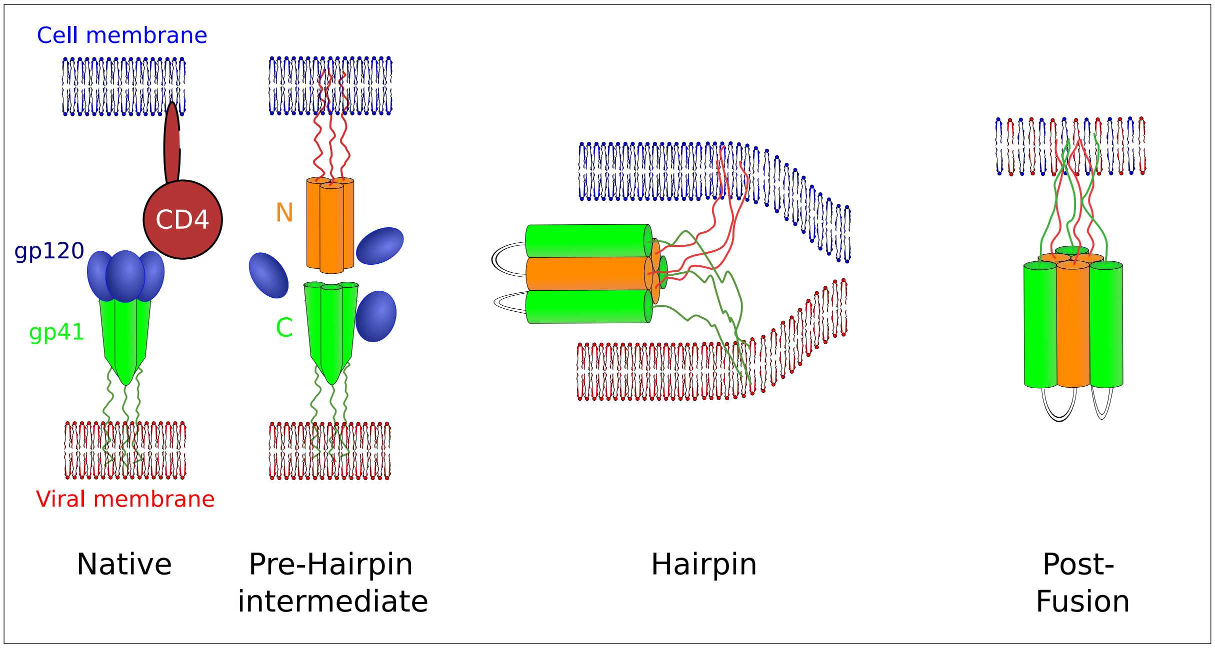


Fig. 4.2: Model of HIV-envelope-mediated membrane fusion. Upon binding to cell surface receptor CD4, gp120 undergoes conformational changes that allow gp41 fusion peptide insertion into the target membrane and the formation of a pre-hairpin structure. Upon gp120 dissociation gp41 transforms into a six-helix boundle. The formation of the six-helix bundle promotes complete fusion, in which the fusion peptide and the transmembrane segment of gp41 lie parallel on a contiguous bilayer [2–4].

supports the following mechanism (Fig. 4.2). The trimeric envelope glycoprotein spike contains gp41 in a prefusogenic conformation. The membrane fusion is initiated by conformational changes of gp120/gp41 when gp120 binds to CD4 and subsequently to the co-receptor. A transient gp41 species or the so-called pre-hairpin intermediate is formed by exposure of the fusion-peptide region and concurrent formation of the N-terminal coiled-coil trimer. The association of the C-terminal heptad-repeat region with the N-terminal coiled-coil resolves the prehairpin intermediate into the fusion-active hairpin structure and leads to apposition of viral and cellular membranes [4].

According to this model, the energy released by the formation of the trimer-of-hairpins is used to overcome the energy barrier for fusion of two lipid bilayers, which is energetically unfavorable. In other words, the refolding transition of gp41 into its fusogenic conformation and the process of membrane fusion are mechanistically and thermodynamically coupled. Therefore, both the conformational stability of the trimer-of-hairpins structure and the rate of its formation may play a critical role in driving membrane fusion. Indeed, there is evidence that destabilizing mutations within the six-helix bundle correlate with phenotypic changes.

Previous work from this laboratory (see [7] and Section 5.1 of the Appendix) demonstrated that engineered constructs of the HIV and SIV gp41, which contain the characteristic features of the fusogenic state of gp41 (six-helix bundle; see Figure 4.1B) are amenable to biophysical characterization. It was shown that point mutations can have dramatic effects on the stability and folding kinetics. The present work continues the efforts to understand the structural determinants of the stability and folding of the trimer-of-hairpins.

4.2 EXPERIMENTAL PROCEDURES

The construction of the expression vector, the expression and purification protocols of the SIV gp41 protein have been described [7, 8]. Additionly, one step was introduced in protein purification: after dialysis against 4 % acetic acid, the dyalisate was frozen and subsequently centrifugated at 20000 g for one hour to remove remaining cell debris prior to HPLC purification. Site directed mutagenesis was performed by standard protocols using the Quick Change kits from Qiagene. The molecular mass of the mutants was verified by mass spectrometry. Since no mutation involves replacement of an aromatic residue, protein concentration was determined using $\varepsilon_{280} = 17070\ \mathrm{M}^{-1}\ \mathrm{cm}^{-1}$ [9]. In the text, concentrations are given in monomer equivalents. All experiments were performed at 25 ℃ in PBS buffer (8.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4 (pH 7.0), 137 mM NaCl, 2.7 mM KCl; I=164 mM). The pH of buffers containing urea was adjusted after addition of the denaturant. Urea concentrations were determined by measuring the refraction index.

Equilibrium unfolding experiments, folding/unfolding kinetic experiments, and primary data analysis were performed using instrumentation and mathematical formalism as described in Chapter 2. Φ-values were calculated as:

$$\Phi = \frac{lnk_f^{WT} - lnk_f^{MUT}}{lnK_{eq}^{WT} - lnK_{eq}^{MUT}} \tag{4.1}$$

$\mathrm{K}_{\mathrm{eq}}^{\mathrm{WT}}$ and $\mathrm{K}_{\mathrm{eq}}^{\mathrm{MUT}}$ are the equilibrium unfolding constants of the wild type and the mutant proteins, respectively. In this case they are defined as $\mathrm{K}_{\mathrm{eq}} = \frac{[\mathrm{T}]}{[\mathrm{M}]^3}$. $\mathrm{k}_{\mathrm{f}}^{\mathrm{WT}}$ and $\mathrm{k}_{\mathrm{f}}^{\mathrm{MUT}}$ are the corresponding folding rate constants. Eq 4.1 is generally valid for any solvent condition. To calculate Φ in aqueous buffer, equilibrium and kinetic data collected in denaturant-containing solutions were extrapolated to obtain K_{eq} and k_{f} according to:

$$lnK_{eq} = lnK_{eq}(D) + \frac{m_{eq}}{RT}[D] \tag{4.2}$$

$$lnk_f = lnk_f(D) + m_f[D] \tag{4.3}$$

$\mathrm{K}_{\mathrm{eq}}(\mathrm{D})$ and $\mathrm{k}_{\mathrm{f}}(\mathrm{D})$ are the equilibrium constant and the folding rate constant at a given denaturant concentration, [D]. ($\mathrm{R} = 8.314\ \mathrm{J\ K}^{-1}\ \mathrm{mol}^{-1}$ is the gas constant and T is the absolute temperature). The terms m_{eq} (in units of $\mathrm{kJ\ mol}^{-1}\ \mathrm{M}^{-1}$) and m_{f} (in units of M^{-1}) describe the dependence of the logarithm of the equilibrium constant and of the logarithm of the folding rate constant on [D] in the general form $\mathrm{m} = \frac{\mathrm{dlnk}}{\mathrm{d[D]}}$. Likewise, the unfolding rate constant was calculated as:

$$lnk_u = lnk_u(D) + m_u[D] \tag{4.4}$$

Sedimentation equilibrium studies were performed on a Beckman XL-A analytical ultracentrifuge with an An 50-Ti rotor at 25 ℃. Protein solutions were dialyzed overnight against PBS, loaded at initial concentrations of 10 μM and analyzed at rotor speed of 25000 rpm. Data sets (five per protein) were fitted simultaneously to a single-species model with the program NONLIN [10]. The protein partial specific volume and solvent density were calculated were calculated using Sednterp [11]. Molecular masses were all within 5% of those calculated for an ideal trimer, In no case a systematic deviation from the statistical distribution of residuals was detected.

4.3 RESULTS AND DISCUSSION

4.3.1 Experimental design

Prerequisites for a reliable mutational analysis of a protein can be summarized as follows. *(i)* Neither wild type protein, nor any mutant is prone to aggregation in the concentration range suitable for biophysical experiments. *(ii)* The protein has the appropriate stability to withstand destabilizing mutations. *(iii)* There are accurate and reproducible methods available for measuring folding/unfolding kinetics and equilibrium transitions. *(iv)* Spectroscopic probes are available. *(v)* Empirical (visual) or force filed-based structural analysis suggests the existence of sites, where non-disruptive mutations can be introduced. *(vi)* For a straightforward characterization of the transition state (TS) the protein folding/unfolding should pertain as close as possible to a simple two-state reaction. Considering conditions *(iii)*-*(vi)*, both SIV-based and HIV-based constructs are suitable for studies. We opted for the SIV protein (and not for a HIV-based construct) since SIV gp41 has higher solubility and better in-solution behavior, as the available information documents [7, 8]. However, the wild type protein (strain $\mathrm{SIV_{MAC}}$), which we have previously characterized, is not very stable (mid-point of urea denaturation was 1.1 M with 5 μM protein; [7]), meaning that condition *(ii)* is not fulfilled. To avoid this problem and to be able to properly measure the folding kinetics in a wider concentration range of denaturants we choose the T28I protein as a pseudo wild type, since its mid-point of urea denaturation is shifted by 2.9 M to higher concentration in comparison to the genuine $\mathrm{SIV_{MAC}}$ wild type protein. Like the genuine wild type, this mutation was studied earlier in great detail in respect to its thermodynamic [7] and kinetic [12] properties. The free energy obtained from equilibrium measurements was the same as the one calculated from folding and unfolding kinetics. The m-value from equilibrium and kinetic experiments,

m_{eq} and $m_{kin} = RT(\mid m_f \mid + \mid m_u \mid)$, respectively, were the same within error. Finally the Tanford β-values (β_T), which describes the position of TS along the reaction coordinate, are reasonably close (0.85 for the wild-type, compared to 0.79 for the pseudo wild-type). For simplicity, we henceforth refer to the T28I pseudo wild type as to WT.

The available X-ray and NMR 3D structures of SIV/HIV gp41 constructs reveal a conservative folding pattern. Our mutational strategy was the following. To probe the role of interactions stabilizing the internal three-stranded coiled coil we converted all **a** and **d** positions of the four-heptad long 3,4-repeat of the N-terminal helices to alanine. In the course of preliminary experiments, however, it turned out that alanine replacing threonine in **a** position 24 and alanine replacing isoleucine in **d** position 28 prevent folding. The protein was expressed in high yields, was soluble, yet no helical signal was observed up to 200 μM concentration. To overcome the problem we designed mutations T24I and I28T. The choice in selecting these replacements was guided by the observation that these positions are conservatively occupied by either threonine or isoleucine in naturally occurring SIV and HIV strains [7]. Further, in the C-terminal helices we converted to alanine hydrophobic residues that contribute to the hydrophobic core by making contacts exclusively with hydrophobic groups located in the internal coiled coil trimer. Altogether thirteen mutants were characterized. Part of them is non-canonical (polar versus non-polar side chains). The rest create relative big packing defects (Ile/Leu versus Ala). We consciously did not consider Ile/Leu→Val substitutions. It is well known that the

Tab. 4.1: AUC characterization of wild type and variants of the SIV-gp-41 six-helix bundle protein.[a]

Variant	Theoretical mass (Da)	Observed mass (Da)
WT	23841	23900
Q7A	23670	22900
L10A	23715	23200
V14A	23757	23000
Q17A	23670	23400
L21A	23715	22900
I24T	23805	23900
T28I	23877	23500
L31A	23715	23500
V48A	23757	23400
L51A	23715	23000
I55A	23715	23200
L58A	23715	23100
L59A	23715	23000

[a] All experiments were performed at 25 °C in PBS buffer, pH 7.0. Protein concentration was ca. 10 μM. Rotor speed was 25000 rpm. Estimated error on observed mass is ± 100 Da.

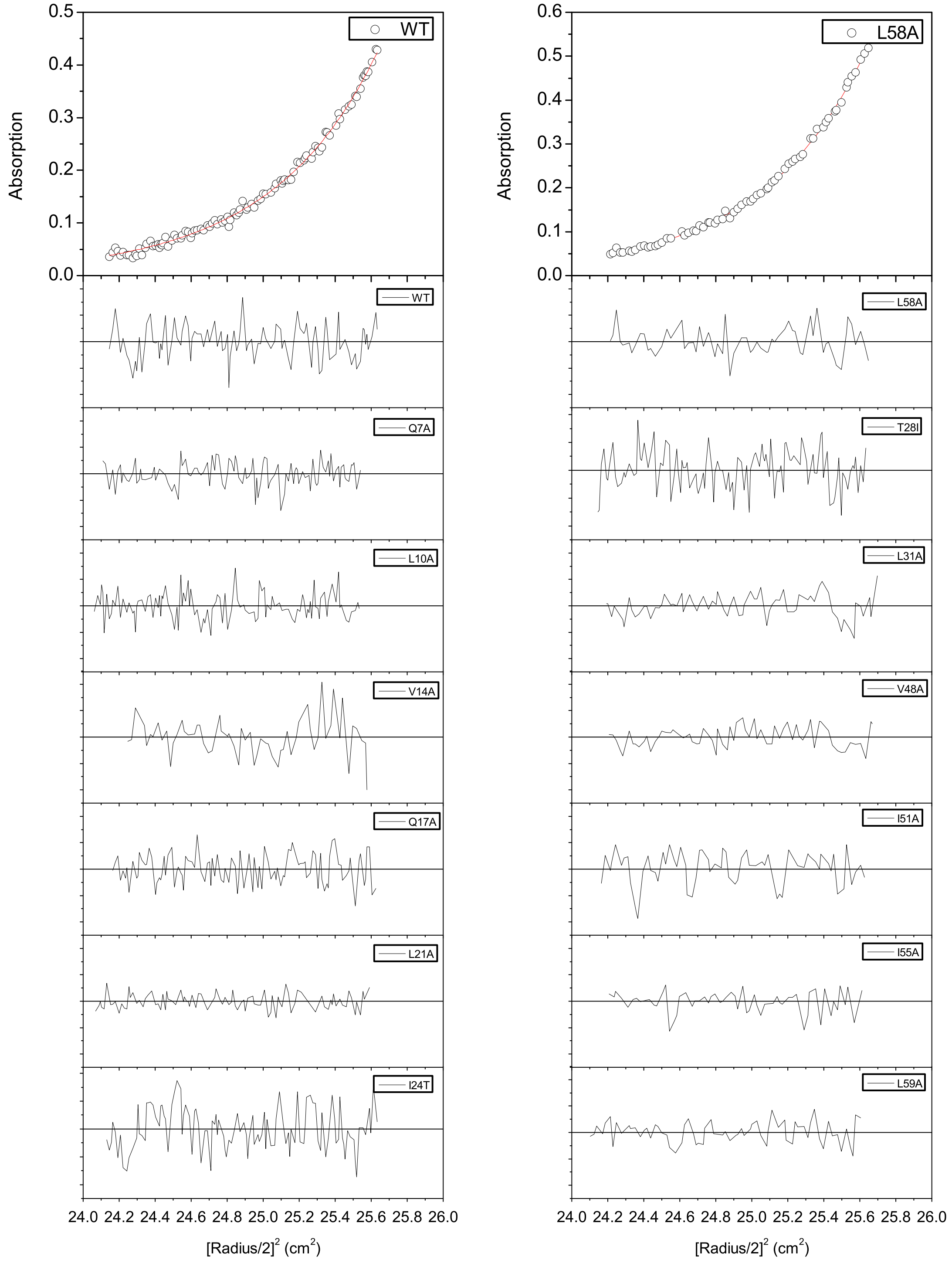


Fig. 4.3: Analytical ultracentrifugation data were collected at 20 °C in PBS. In all cases the protein concentration was ca. 10 uM. The data were fitted to a ideal trimeric species, and all fitted values were within 5% of theoretically calculated values for given variant. The toppanels show the results obtainded with the most stable (WT) and most unstable (L58A) proteins. For all other mutants, only the distribution of residuals is shown.

strandneness of coiled coils critically depends on the nature of hydrophobic side chains in **a** and **d** positions. Indeed, all mutant proteins are trimers (Fig. 4.3 and Table 4.1).

On the other hand, in a parallel, in-register homotrimer, removal of even one methylene group would produce a large cavity in an **a** or **d** layer. Nevertheless, the mutation sites are spread over the length of the protein and are potentially useful to understand the energetic and kinetic role of wild type interactions by "quasi-Φ-value analysis".

4.3.2 Equilibrium unfolding

The stability of all mutants was assessed from urea-induced unfolding at 25 C. Fig. 4.4 represents the fraction of unfolded protein, f_U, as function of the denaturant concentration, [D].

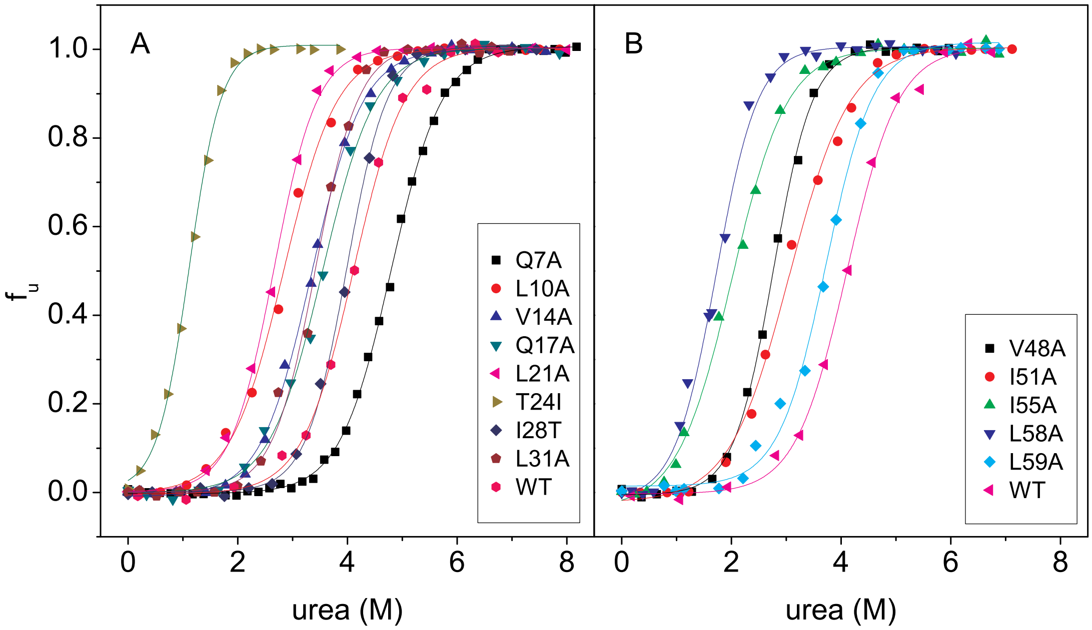


Fig. 4.4: Equilibrium isothermal unfolding of gp41 variants by urea. Shown are normalized unfolding curves recorded in PBS at 25 °C. The symbols are the experimental data and the continuous lines are calculated as described in chapter 2.2.

As usual, the fraction unfolded protein was calculated from the raw experimental traces (θ_{222} at various denaturant concentrations) assuming that the linear changes of θ_{222} at low and high urea concentrations represent the intrinsic ellipticity changes of the folded (trimeric) and unfolded (monomeric) ensembles. We noted that these $\frac{d\theta_{222}}{d[D]}$ functions for different mutants scatter significantly but did not detected any systematic trend (i.e. N-terminal versus C-terminal substitutions, or polar versus non-polar WT residues, or larger versus smaller WT residues). The meaning of the pre-transitional and post-transitional $\frac{d\theta_{222}}{d[D]}$] changes has been debated (for example in ref. [13]), yet no consensus has been reached. As seen in Fig. 4.4, within this approximation (the protein is fully folded at the on-set of the unfolding transition and is fully unfolded after the unfolding transition), the unfolding curves exhibit no "steps" as to indicate the presence of intermediates differing in helical content. Following the (assumed) two-state model and the linear extrapolation model (LEM) the unfolding free energy changes in aqueous buffer, ΔG_{eq}, and the m_{eq} coefficients were calculated and are listed in Table 4.2. Note that the mid-points of urea unfolding, $[D]_{1/2}$, which are also given in Table 4.2 cannot be directly compared since the experiments were performed at different protein concentrations. Nevertheless, with the exception of Q7A and T24I variants, both ΔG_{eq} and $[D]_{1/2}$ values are lower than those measured for the WT protein, thus indicating significant destabilization effect of the corresponding mutations (Table 4.2).

The energetic neutrality of the Q7A and T24I substitutions is intriguing, since in both cases a polar function is eliminated. Comparison with the highly destabilizing I28T mutation (which is the most destabilizing in the set!) and the modestly destabilizing Q17A mutation reveals that in the formally identical **a** positions of the first and fourth heptad (7 and 28, respectively) non-polar interactions confer stability, while in the **d** positions of the second and third heptad (17 and 24, respectively) polar interactions are important. The energetic non-equivalence of **a** and **d** positions in coiled coils has been demonstrated [14, 15].

Hydrophobic side chains in the middle of the central coiled coil (L10, V14, L21) stabilize the trimer to approximately the same extent. Differently, there is a clear trend that mutating out side chains that link the C-terminal helices to the central coiled coil by hydrophobic contacts is more destabilizing the closer the corresponding sites are to the middle of the six-helix bundle (compare the increasing effect of V48A, L51A, I55A and L58A). The small destabilizing effect of the L59A substitution is suprising. Simple structural analysis (of highly related X-ray structures of SIV and HIV constructs) rationalizes the large difference between the L58A and L59A mutations by the fact that, although both side chains are completely buried, the hydrophobic packing is much more optimized for L58 than for L59. This is only a tentative explanation since both L58A and L59A show the largest deviations from the 2-state model (see below). Excluding L59A, among the conservative hydrophobic-to-alanine mutations, variants L31A and V48A have the smallest effect on stability. This probably reflects the fact that these two sites are closest to the loop region and experience larger structural fluctuations (less well optimized hydrophobic complementarity) than the rest of the probed sites.

The scatter of the m_{eq}-values, 11.9 ± 1.7 kJ mol^{-1} M^{-1} (mean ± SD), exceeds the experimental error ($m_{eq} = 0.5$ kJ mol^{-1} M^{-1}), but all m_{eq}-values are in very good agreement with statistical analysis of the protein database, which predicts an urea denaturation meq-value of 11.7 ± 0.9 kJ mol^{-1} M^{-1} for a protein consisting of 210 residues [16, 17]. Hence, the measured m_{eq} support the trimeric nature of all variants, the expected values being 7.9 kJ mol^{-1} M^{-1} and 15.5 kJ mol^{-1} M^{-1} for a dimer and a tetramer, respectively, and indicate, in general, low population of possible intermediates. (However, see later).

Tab. 4.2: Thermodynamic and kinetic parameters describing the stability of SIV gp 41 six-helix bundle.[a]

Variant[b]	$U_{1/2}^{eq}$ (M)	m_{eq}^{calc} (kJ mol^{-1} M^{-1})	ΔG_{eq} (kJ mol^{-1})	ΔG_{kin} [c] (kJ mol^{-1})	k_f (M^{-2} s^{-1})	k_u (s^{-1})	m_f (M^{-1})	m_u (M^{-1})	m_{kin} [d] (kJ mol^{-1} M^{-1})	β_T [e]	Φ_f [f]
WT	**4.1**	**13.5**	**112**	**115**	**1.6×10^{14}**	**1.1×10^{-6}**	**-3.8**	**1.0**	**11.9**	**0.79**	–
Q7A	4.8	11.0	111	94	1.8×10^{14}	2.5×10^{-3}	-2.7	0.4	7.6	0.88	(1.29)
L10A	**2.8**	**10.6**	**89**	**86**	**1.3×10^{12}**	**1.2×10^{-3}**	**-3.3**	**0.8**	**10.2**	**0.8**	**0.51**
V14A	**3.2**	**11.8**	**96**	**94**	**1.7×10^{12}**	**4.8×10^{-5}**	**-2.7**	**0.8**	**8.7**	**0.78**	**0.69**
Q17A	**3.5**	**10.4**	**95**	**95**	**1.4×10^{14}**	**2.9×10^{-3}**	**-2.0**	**1.2**	**7.9**	**0.62**	**0.02**
L21A	2.7	13.1	91	78	1.4×10^{11}	3.7×10^{-3}	-3.5	0.4	9.4	0.91	(0.82)
T24I	4.0	12.9	111	91	5.6×10^{12}	5.4×10^{-4}	-2.1	1.1	7.8	0.66	(1.50)
I28T	**1.1**	**16.1**	**75**	**76**	**1.2×10^{11}**	**5.8×10^{-3}**	**-5.5**	**1.3**	**16.8**	**0.81**	**0.48**
L31A	**3.4**	**12.6**	**102**	**95**	**6.1×10^{12}**	**1.4×10^{-4}**	**-2.8**	**1.3**	**10.2**	**0.69**	**0.80**
V48A	**2.8**	**14.3**	**98**	**97**	**2.3×10^{12}**	**2.1×10^{-5}**	**-2.7**	**1.3**	**10.1**	**0.67**	**0.73**
I51A	**3.1**	**9.6**	**88**	**94**	**5.2×10^{12}**	**1.6×10^{-4}**	**-4.3**	**1.0**	**13.2**	**0.81**	**0.36**
I55A	**2.0**	**11.0**	**83**	**91**	**4.9×10^{11}**	**6.2×10^{-5}**	**-2.5**	**1.5**	**9.9**	**0.62**	**0.50**
L58A	1.7	14.0	77	114	3.2×10^{14}	3.6×10^{-6}	-5.1	1.3	15.8	0.79	(-0.05)
L59A	3.5	12.5	104	80	1.4×10^{13}	1.4×10^{-1}	-4.1	1.0	12.7	0.81	(0.71)

[a] All experiments were performed at 25 °C ain PBS buffer, pH 7.0. Protein concentration was 5-10 μM (monomer equivalents). The estimated errorrs are: 0.1 M in $U_{1/2}^{eq}$; 0.5 kJ mol^{-1} M^{-1} in m_{eq}^{calc}; 7-10 kJ mol^{-1} in ΔG_{eq}; 7-10 kJ mol^{-1} in ΔG_{kin}; ∼5 % in k_f; ∼10 % in k_u; 0.1 M^{-1} in m_f; 0.1 M^{-1} in m_u; 0.5 kJ mol^{-1} M^{-1} in m_{kin}; 0.03 in β_T and 0.05 in Φ_f.

[b] Bold face is used for variants exhibiting two-state behavior according to the following criterion: $\Delta\Delta G = \mid \Delta G_{eq} - \Delta G_{kin} \mid < 10$ kJ mol^{-1}.

[c] Calculated according to $\Delta G_{kin} = -RT\ln\frac{k_u}{k_f}$

[d] Calculated according to $m_{kin} = RT(|m_f| + |m_u|)$

[e] Calculated according to $\beta_T = 1 - \frac{m_u}{|m_f|+|m_u|}$

[f] Calculated according to $\Phi = \dfrac{\ln k_f^{WT} - \ln k_f^{MUT}}{\ln K_{eq}^{WT} - \ln K_{eq}^{MUT}}$

4.3.3 Folding/unfolding kinetics

Refolding and unfolding experiments were performed by following the time course of θ_{222} after rapid transfer of denatured protein to folding conditions, or of native protein to denaturing conditions. Representative kinetic traces are shown in Fig. 4.5.

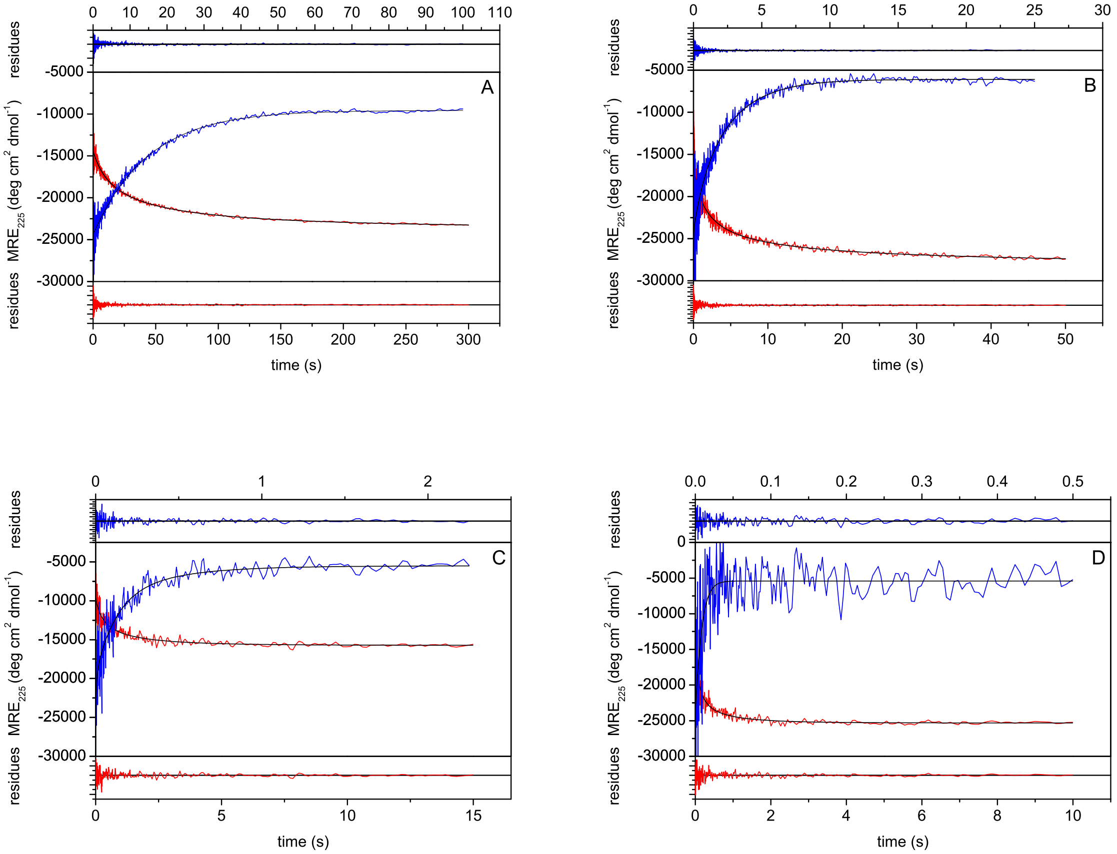


Fig. 4.5: Representative kinetic traces for folding and unfolding of gp41. As an illustration, the fastest and the slowest reactions measured for L10A (panels A and B) and L59A (panels C and D) as example for a 2SM and non-2SM variant, respectively. In all panels, folding traces are shown in red, and unfolding traces are shown in blue. (Panel A) Folding/unfolding of L10A in 1.48 M urea (12.3 μM L10A) and in 4.8 M urea (6.8 μM L10A), respectively. (Panel B) Folding/unfolding of L10A in 0.75 M urea (12.3 μM L10A) and in 7.15 M urea (6.8 μM L10A), respectively. (Panel C) Folding/unfolding of L59A in 1.86 M urea (10.9 μM L59A) and in 2.59 M urea (6.3 μM L59A), respectively. (Panel D) Folding/unfolding of L59A in 0.59 M urea (10.9 μM L59A) and in 6.47 M urea (6.3 μM L59A), respectively.

As exemplified in this figure, all collected traces can be perfectly described by numerical integration of the following equations, which describe the appearance or disappearance of the native trimer in refolding experiments or unfolding experiments, respectively:

$$\frac{d[M]}{dt} = -3k_{on}[M]^3 + 3k_{off}[T] \tag{4.5}$$

$$\frac{d[T]}{dt} = k_{on}[M]^3 - k_{off}[T] \tag{4.6}$$

The success of the analysis favors the simplest two-state model:

$$M + M + M \rightleftharpoons T \tag{4.7}$$

according to which only folded trimers and unfolded monomers are significantly populated along the reaction coordinate. As discussed previously [12, 18, 19], the presence of dimeric or trimeric intermediates cannot be excluded, yet they must be spectrosopically silent, i.e. they exhibit ellipticity signature which is indistinguishable within error from those of the folded and unfolded protein. For further analysis we consider only kinetic experiments, which conform to the following requirements. First, the observed kinetic amplitude ($\Delta\theta_{225} = \theta_{225,(\mathrm{t=0})} - \theta_{225,(\mathrm{t=\infty})}$ was at least 90% of the amplitude detected at equilibrium at each tested urea concentration. Second, the observed time window was shorter than 200 s to exclude baseline-drift artifacts. (So far we have no evidence for slow folding phases. The protein contains no prolines to introduce kinetic complications. Neither is there evidence for formation of burst-phase helical intermediates.) These two conditions somewhat reduce the range of urea concentrations at which folding/unfolding kinetics was subjected to analysis, but provide a rigorous basis for analysis of the major folding/unfolding event. (It is a priori clear that minor conformational changes are not amenable to CD experiments.)

Following LEM, the microscopic rate constants, $\mathrm{k_f}$ and $\mathrm{k_u}$, were calculated by extrapolation of the $\ln(\mathrm{k_f})$ or $\ln(\mathrm{k_f})$ versus [D] functions to [D] = 0 according to eqs 4.3 and 4.4. The Chevron plots, which are all linear in the studied range of urea concentrations, are shown in Fig. 4.6. The derived rate constants are listed in Table 4.2 along with the corresponding $\mathrm{m_f}$ and $\mathrm{m_u}$ coefficients. Most of the mutants fold slower and unfold faster than the WT protein (Fig. 4.7). In terms of folding rates, mutations Q7A, Q17 and L58A are kinetically inert, their relative folding rates, $\mathrm{k_{f,rel}} = \frac{\mathrm{k^{MUT}_{f,rel}}}{\mathrm{k^{WT}_{f,rel}}}$ being between 0.5 and 2. For five mutants, V14A, T24I, L31A, V48A, I51A, and L59A, $\mathrm{k_{f,rel}} > 0.01$. Slowest folding ($\mathrm{k_{f,rel}} < 0.01$) show variants L10A, L21A, I28T, and I55A. In terms of unfolding, all variants are destabilized. The relative unfolding rates, $\mathrm{k_{u,rel}} = \frac{\mathrm{k^{MUT}_{u,rel}}}{\mathrm{k^{WT}_{u,rel}}}$, rank the mutations in three (also arbitrarily defined) categories: *(i)* lowly destabilized ($10 < \mathrm{k_{u,rel}} < 100$; V14A, V48A, I55A), *(ii)* modestly destabilized ($100 < \mathrm{k_{u,rel}} < 1000$; T24I, L31A, I51A), and *(iii)* highly destabilized ($\mathrm{k_{u,rel}} > 1000$; Q7A, L10A, Q17A, L21A, I28A). The clear exceptions are variants L58A ($\mathrm{k_{u,rel}} < 5$) and L59A ($\mathrm{k_{u,rel}} > 1 \times 10^5$). The former unfolds like the WT protein; the latter unfolds dramatically faster. As already mentioned, and as it will be discussed bellow, the assumptions implicit to the two-state model are not met by these variants, thus making it difficult to rationalize such a tremendous difference in kinetic stability caused by formally identical mutations. However, we note that these sequentially adjacent side chains make hydrophobic links to different inner helices. As a tentative explanation, we suggest that the L58A/L59A “hydrophobic fork“ located

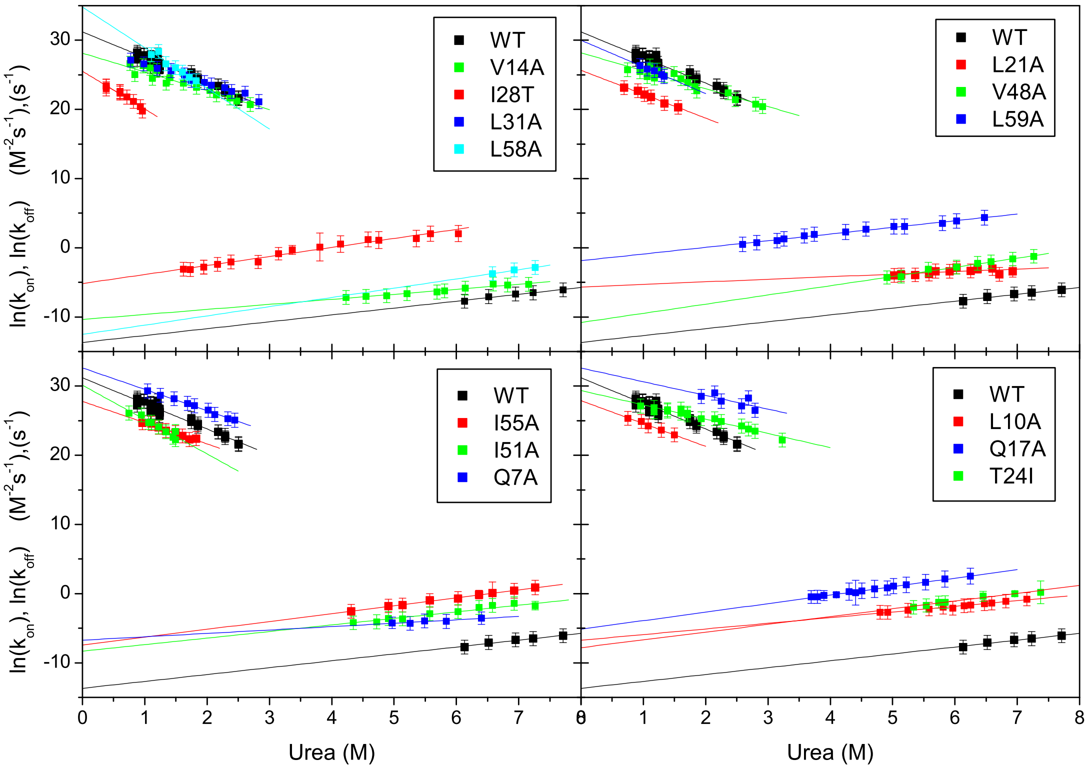


Fig. 4.6: Denaturant dependence of the folding and unfolding kinetics (chevron plots) of wild-type gp41 (WT) and variants. The Chevrons are ploted arbitrary to alow comparisons as easy as possible. Data with the WT are preseuted in each panel(black squares). All plots are scaled identically.

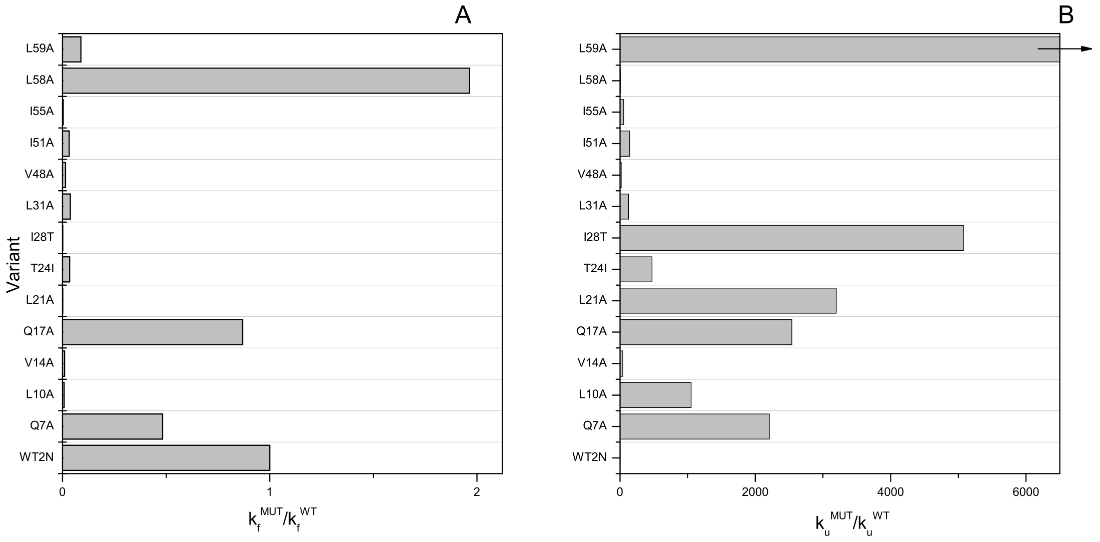


Fig. 4.7: Relative changes of the rate constants of refolding (k_f; panel A) and unfolding (k_u; panel B) measured for variants of the SIV gp41 six-helix bundle construct. The arrow in panel B indicates that $k_u^{L59A}/k_u^{WT} > 7000$.

approximately half way from the ends of the six-helix bundle is a hot spot in stabilizing the native state, the hydrophobic packing of L59 being less well optimized (see above). Removal of three methylene groups by mutation to alanine introduces a packing defect. In this particular site, the loss of hydrophobic interactions involving L59 cannot be compensated by neighboring interactions, and, as the consequence, thermal fluctuations overcome some critical threshold. As seen in Fig. 4.10, the apparent Φ-values of L58 and L59 are very low and rather high, respectively. Notwithstanding all reservations in interpreting this difference (both variants are non-two-state folders), proper set of interactions made by L59 seems to be crucial.

Again, the changes in the folding and unfolding rates caused by non-canonical polar-versus-nonpolar substitutions deserve attention. It appears that glutamine-to-alanine mutations are balanced in a similar way by compensating changes (decrease) in $\mathrm{k_f}$ and $\mathrm{k_u}$. In heptades 1 and 4 of the N-terminal coiled coil, nonpolar side chains in a positions seems to speed up folding more than slow down unfolding (compare the changes in $\mathrm{k_f}$ and $\mathrm{k_u}$ of Q7A and I28T relative to WT). In contrast, polar functions in positions d of heptades 2 and 3 are beneficiary in terms of both folding and unfolding, in agreement with the conclusions drawn from analysis of $\Delta \mathrm{G_{eq}}$ (see above).

4.3.4 Two-state versus non two-state folding/unfolding mechanism

In the framework of classical mutational (Φ-value) analysis of the folding/unfolding mechanism, the definition of criteria ruling the acceptance or rejection of the two-state approximation is a crucial step. It is clear that *(i)* equilibrium, quasi-sigmoidal unfolding curves,*(ii)* kinetic, quasi-exponential traces, and *(iii)* Chevron plots, in the absence of abrupt signal changes, can be formally discussed as describing a two-state process. A stronger criterion is the comparison of the free energy changes derived form eqiulibrium and kinetic experiments,$\Delta \mathrm{G_{eq}} = -\mathrm{RTlnK_{eq}}$ and $\Delta \mathrm{G_{kin}} = -\mathrm{RTlnK_{kin}} = -\mathrm{RTln}\frac{\mathrm{k_f}}{\mathrm{k_u}}$, respectively. Fig. 4.8 illustrates that $\Delta \mathrm{G_{eq}}$ and $\Delta \mathrm{G_{kin}}$ are not correlated overall ($\mathrm{R}^2 = 0.04$), in sharp contrast to the definition of two-state behaviour. However, the experimental uncertainties should be considered. We conservatively estimate the mean ΔG error in the set as being 7-10 kJ mol^{-1}. According to this criterion, nine proteins can be classified as two-state folders (WT, L10A, V14A, Q17A, I28T, L31A, V48A, I51A, and I55A). The $\Delta \mathrm{G_{kin}}$-versus-$\Delta \mathrm{G_{eq}}$ correlation coefficient for these proteins is $\mathrm{R}^2 = 0.951$ (Fig. 4.8).

Another test for the fulfilment of the two-state mechanism is the comparison of the Brnsted coefficients for the folding and unfolding reactions [20]. The Brønsted coefficient is defined in the following way:

$$\beta = \frac{dlnk}{dlnK} \tag{4.8}$$

where k is a rate constant and K is the genuine equilibrium constant (derived from ex-

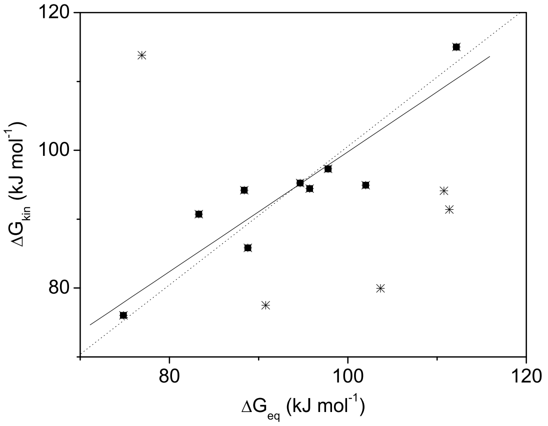


Fig. 4.8: Correlation between the unfolding free energy changes obtained from equilibrium (ΔG_{eq}) and kinetic (ΔG_{kin}) experiments. $\Delta G'$s were calculated assuming a two-state model of folding/unfolding ($\Delta G_{eq} = -RT\ln K_{eq}$; $\Delta G_{kin} = -RT\ln(k_u/kf)$). The squares represent data fitting the following criterion:$\mid \Delta G_{eq} - \Delta G_{kin} \mid < 10$ kJ mol^{-1}. The solid line is the correlation line describing these data only ($R^2 = 0.951$). The dashed line represents the ideal expected correlation between ΔG_{eq} and ΔG_{kin} ($R^2 = 1$).

periment(s) performed at equilibrium conditions, and not derived from the k_u/k_f ratio). In protein folding, the two-state model is valid if the Brønsted coefficient for the folding and unfolding reactions, β_f and β_u, respectively, add up to unity, i. e. :

$$\mid \beta_f \mid + \mid \beta_u \mid = 1 \tag{4.9}$$

In the studied set of mutations, both β_f and β_u are close to zero (Fig. 4.9). However, for proteins that have been selected as two-state folders, based on the ΔG_{kin}-versus-ΔG_{eq} correlation,$\mid \beta_f \mid + \mid \beta_u \mid = 0.9 \pm 0.2$.

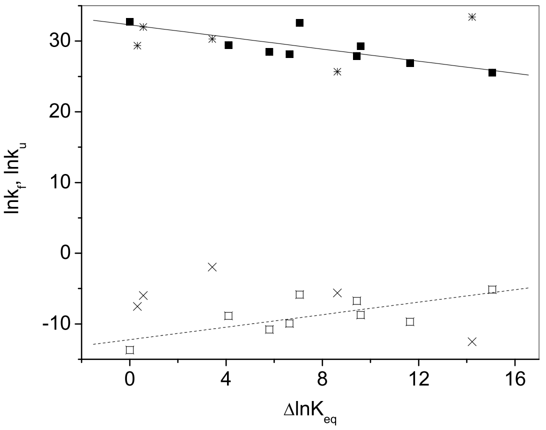


Fig. 4.9: Brønsted plot for folding and unfolding of SIV gp41 trimer-of-hairpins. The solid symbols and asterisks represent the complete set of $\ln k_f$ data. Likewise, the open symbols and crosses represent the complete set of $\ln k_u$ data. $\Delta\ln K = \ln(K^{MUT}/K^{WT})$ is the change of the unfolding constant calculated from equilibrium data. The squares highlight $\ln k_f$ (solid) and $\ln k_u$ (open) for proteins for which $0 <\mid \Delta G_{eq} - \Delta G_{kin} \mid < 10$ kJ mol^{-1} (see the text for details). The lines are best linear fits to the folding and unfolding data (continuous and dashed lines, respectively) considering only these variants. The values of $\mid$ slope $\mid = \beta = d\ln k/d\Delta K$ representing the Brnsted coefficients are 0.45 ± 0.13 (folding) and 0.44 ± 0.16 (unfolding).

Finally, we examine the correspondence between the m_{eq} and $m_{kin} = RT(\mid m_f \mid + \mid m_u \mid)$ coefficients as a test for the validity of the two-state model. For a two-state transition, $m_{eq} = m_{kin}$. The two sets of m-values do not differ statistically ($m_{eq} = 11.9 \pm 1.7$ and $m_{kin} = 10.9 \pm 2.8$ kJ mol^{-1} M^{-1}; mean ± SD). However, all but three mutations, L10A, I55A, and L59A, exhibit $m_{eq} - m_{kin}$ differences larger than the estimated experimental error. (Note that the L59A protein doesnt fit the ΔG_{kin}-versus-ΔG_{eq} criterion as a two-state folder). On a closer look, the difference between $\mathrm{m_{eq}}$ and $\mathrm{m_{kin}}$ stems from a larger variation in $\mathrm{m_f}$ (Table 4.2). Furthermore, there are no cross-correlations between $\mathrm{m_{eq}}$ and $\mathrm{m_{kin}}$ to $\Delta\mathrm{G_{eq}}$ and $\Delta\mathrm{G_{kin}}$ (in absolute values and in relative changes). The reason for the discrepancies between $\Delta\mathrm{G_{eq}}$ and $\Delta\mathrm{G_{kin}}$ is not known.

Previously, the WT, T24I and I28T [7, 12], as well as two HIV gp41 proteins [21] have been classified as two-state folders based on equilibrium information only (analysis of calorimetric experiments and comparisons between thermal and isothermal unfolding data). The kinetic data collected in this work add a further level of analysis and demonstrate that some variants clearly deviate from the two-state behaviour, and most proteins cannot be classified as two-state folders without some reservations. Nevertheless, in view of the experimental errors, nine proteins fold/unfold with a reasonable resemblance to a two-state process, as to provide hints about the gross structure of the transition state.

4.3.5 Properties of the transition state

The structure of the rate-determining, high-energy state, which proteins transverse on the folding/unfolding coordinate, has attracted considerable attention in recent years, since it provides important clues about the sequence of structural events along the folding/unfolding pathway (i.e. of the folding/unfolding mechanism). The identification of the set of interactions realised in the transition state is only possible by mutation in the framework of the Φ-value analysis [20, 22]. Fig. 4.10 shows the Φ-values calculated according to eq 4.1. In the set of quasi-two-state folders, L10A, I28T, I51A, and I55A exhibit $\Phi \leq 0.5$. For three mutants, V14A, L31A, and I55A, $\Phi \geq 0.5$. In the usual interpretation, partial Φ-values ($0.3 \leq \Phi \leq 0.7$) are difficult to interpret [23–25]. Such Φ-values indicate either partial formation of native-like interactions involving the WT side chain in a conformationally homogeneous TS, or the existence of alternative local conformations within the TS ensemble, or even the existence of largely different transition states (parallel folding pathways). The clear outlier is Q17A showing a negligibly small Φ. However, we consider

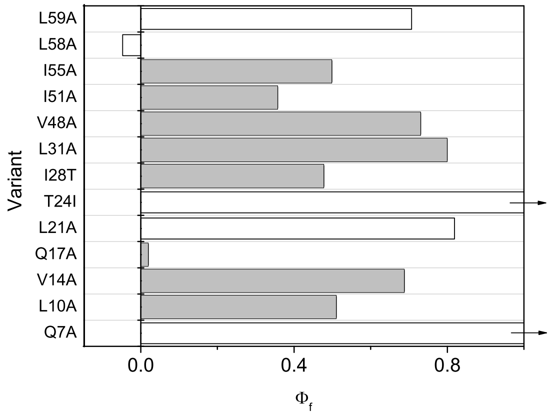


Fig. 4.10: Φ-values characterizing mutations of the SIV gp41 six-helix bundle core. Φ-values were calculated according to eq 4.1. The shaded bars represent the Φ-values for mutations which induce no apparent deviation from the two-state behavior (based on the $0 <| \Delta\mathrm{G_{eq}} - \Delta\mathrm{G_{kin}} | < 10\ \mathrm{kJ\ mol^{-1}}$ criterion (see the text for details). The arrow indicate $\Phi > 0$ (the actaul values are 1.2 and 1.6 for Q7A and T24I, respectively).

this Φ-value as being artificial, since both ΔG_{eq} and k_f are statistically indistinguishable close to the WT parameters.

We are reluctant to discuss the meaning of individual Φ-values in terms of local structure. Nonetheless, the following points are worth noting. *(i)* The mean measured Φ equals 0.50 ± 0.25. *(ii)* The mutations cover three quarters of the length of the elongated six-helix bundle. *(iii)* Similar Φ were measured for sites close to the N-terminus (L10, V14) and to the C-terminus (I28) of the central coiled coil. *(iv)* Similar Φ were measured for side chains participating in the hydrophobic core of the internal coiled coil and side chains linking (by hydrophobic packing) the outer C-terminal helices to the inner N-terminal helices. *(v)* The Φ-values for L31 from the N-terminal helix and V48 from the C-terminal helix are not significantly different. These two side chains are sequentially positioned up-stream and down-stream of the loop region. All these observations strongly suggest that the transition state is compact and the outer and inner helices are oriented in a native-like fashion (antiparallel). Based on the available information TS can be defined as being diffuse in the sense that there is no obvious polarity in the distribution of Φ-values. These conclusions are further supported by considering the apparent Φ-values for non-two-state mutants (Fig. 4.10). The compactness and the diffuse character of TS can also be inferred from the calculated Tanfords β-value. This dimensionless parameter in the two-state approximation is defined as [20, 26]:

$$\beta_T = \frac{\mid m_f \mid}{\mid m_f \mid + \mid m_u \mid} = 1 - \frac{\mid m_u \mid}{\mid m_f \mid + \mid m_u \mid} \tag{4.10}$$

The numerical value of β_T ($0 < \beta_T < 1$) is usually interpreted as an indicator of the position of TS on a reaction coordinate in terms of denaturant-sensitive (i.e. solvent-accessible) surface. β_T close to 1 indicates that TS is nearly as compact as the native state; β_T close to 0 indicates disordered and solvent accessible structure, as in the denatured state. As seen in Table 4.2, all β_T are larger than 0.62 (mean $\beta_T = 0.73 \pm 0.08$ for the two-state variants and 0.76 ± 0.09 for all proteins) and are uniformly distributed along the six-helix bundle. Therefore, TS exhibits a collapsed conformation. Intriguingly, there is a weak, yet obvious trend that both m_f and m_u decrease, the higher the mid-point of equilibrium urea denaturation. This possibly indicates that TS slightly expands at high urea concentrations.

In conclusion, we have probed the consequence of mutating out topologically conserved packing contacts in the hydrophobic core of the six-helix bundle protein construct representing the fusogenic conformation of the SIV gp41 protein. These include both purely non-polar contacts (valine, leucine and isoleucine) and mixed non-polar/polar contacts (glutamine, threonine). It has been proposed that polar side chains buried in the otherwise hydrophobic interior of proteins can be kinetically important determinants [27]. In the present case, there are no obvious (and easily interpretable) trends concerning the role of the probed polar functions in the hydrophobic core of gp41. Also, there are no correlations between the size of the mutated non-polar side chain and the resulting thermodynamic and kinetic changes. Thus, the local geometry and any conformational rearrangements compensating the mutation (mostly cavity-creating) dominate the changes in stability and folding/unfolding rates. To our knowledge, this is the first study aiming at characterization of the transition state of a trimeric protein. The presented data demonstrate that some mutations cause large deviations from the apparent two-state behavior that has been postulated for the wild type six-helix bundle [7, 12]. In fact, only two variants (L10A and I55A) can be strictly classified as two-state folders. This is perhaps not surprising because in a homotrimeric protein three sites are mutated simultaneously. In the specific case of a coiled coil or a parallel, in-register helical bundle, the mutations are located very close to each other, hence introducing a quite large packing defect in a particular spot of the hydrophobic core. Nonetheless, nine proteins obey a weakened two-state criterion. Our "quasi-Φ-value" and β_T-value analyses lead to the conclusion that the six-helix bundle transverses a compact folding transition state, in which native-like structure formation is modestly advanced, yet there are no completely unstructured regions. Therefore, with all necessary reservations, we propose that the highest-energy barrier along the folding pathway is passed in a trimeric state, after the C-terminal half of each monomer chain is fixed in anti-parallel orientation to the surface of the central N-terminal coiled-coil-to-be.

BIBLIOGRAPHY

[1] Trkola, A, Dragic, T, Arthos, J, Binley, J. M, Olson, W. C, Allaway, G. P, Cheng-Mayer, C, Robinson, J, Maddon, P. J, & Moore, J. P. (1996) CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* **384**, 184–187.

[2] Chan, D. C & Kim, P. S. (1998) HIV entry and its inhibition. *Cell* **93**, 681–684.

[3] Gallo, S. A, Finnegan, C. M, Viard, M, Raviv, Y, Dimitrov, A, Rawat, S. S, Puri, A, Durell, S, & Blumenthal, R. (2003) The HIV Env-mediated fusion reaction. *Biochim. Biophys. Acta.* **1614**, 36–50.

[4] LeDuc, D. L & Shin, Y. K. (2000) Insights into a structure-based mechanism of viral membrane fusion. *Biosci. Rep.* **20**, 557–570.

[5] Tan, K, Liu, J, Wang, J, Shen, S, & Lu, M. (1997) Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl. Acad. Sci. U. S .A.* **94**, 12303–12308.

[6] Weissenhorn, W, Dessen, A, Harrison, S. C, Skehel, J. J, & Wiley, D. C. (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387**, 426–430.

[7] Jelesarov, I & Lu, M. (2001) Thermodynamics of trimer-of-hairpins formation by the SIV gp41 envelope protein. *J. Mol. Biol.* **307**, 637–656.

[8] Liu, J, Wang, S, Hoxie, J. A, LaBranche, C. C, & Lu, M. (2002) Mutations that destabilize the gp41 core are determinants for stabilizing the simian immunodeficiency virus-CPmac envelope glycoprotein complex. *J. Biol. Chem.* **277**, 12891–12900.

[9] Edelhoch, H. (1967) Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6**, 1948–1954.

[10] Johnson, M. L, Correia, J. J, Yphantis, D. A, & Halvorson, H. R. (1981) Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* **36**, 575–588.

[11] Laue T., B. Shah, T. R & Pelletier, S. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In: Analytical Ultracentrifugation in Biochemistry and Polymer Science eds. S. Harding, A. R & Horton, J. (The Royal Society of Chemistry, Cambridge, UK), p. 90125.

[12] Marti, D. N, Bjelić, S, Lu, M, Bosshard, H. R, & Jelesarov, I. (2004) Fast folding of the HIV-1 and SIV gp41 six-helix bundles. *J. Mol. Biol.* **336**, 1–8.

[13] Dragan, A. I & Privalov, P. L. (2002) Unfolding of a leucine zipper is not a simple two-state transition. *J. Mol. Biol.* **321**, 891–908.

[14] Jelesarov, I & Bosshard, H. R. (1996) Thermodynamic characterization of the coupled folding and association of heterodimeric coiled coils (leucine zippers). *J. Mol. Biol.* **263**, 344–358.

[15] Zhou, N. E, Kay, C. M, & Hodges, R. S. (1992) Synthetic model proteins: the relative contribution of leucine residues at the nonequivalent positions of the 3-4 hydrophobic repeat to the stability of the two-stranded alpha-helical coiled-coil. *Biochemistry* **31**, 5739–5746.

[16] Geierhaas, C. D, Nickson, A. A, Lindorff-Larsen, K, Clarke, J, & Vendruscolo, M. (2007) BPPred: a Web-based computational tool for predicting biophysical parameters of proteins. *Protein. Sci.* **16**, 125–134.

[17] Myers, J. K, Pace, C. N, & Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein. Sci.*. **4**, 2138–2148.

[18] Bjelić, S, Karshikoff, A, & Jelesarov, I. (2006) Stability and folding/unfolding kinetics of the homotrimeric coiled coil Lpp-56. *Biochemistry* **45**, 8931–8939.

[19] Buchner, J & Kiefhaber, T, eds. (2005) Protein Folding Handbook. (Wiley-VCH).

[20] Fersht, A. (1998) Structure and Mechanism in Protein. Sci.ence: A Guide to Enzyme Catalysis and Protein Folding. (Freeman, W. H., Publishers).

[21] Bjelić, S. (2003) M.Sci. (University of Zurich,Zurich,Schwitzerland).

[22] Fersht, A. R & Sato, S. (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U. S .A.* **101**, 7976–7981.

[23] Neudecker, P, Zarrine-Afsar, A, Davidson, A. R, & Kay, L. E. (2007) Phi-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc. Natl. Acad. Sci. U. S .A.* **104**, 15717–15722.

[24] Weikl, T. R & Dill, K. A. (2007) Transition-states in protein folding kinetics: the structural interpretation of Phi values. *J. Mol. Biol.* **365**, 1578–1586.

[25] Bodenreider, C & Kiefhaber, T. (2005) Interpretation of protein folding psi values. *J. Mol. Biol.* **351**, 393–401.

[26] Pace, N. C & Tanford, C. (1968) Thermodynamics of the unfolding of beta-lactoglobulin A in aqueous urea solutions between 5 and 55 degrees. *Biochemistry* **7**, 198–208.

[27] Waldburger, C. D, Jonsson, T, & Sauer, R. T. (1996) Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl. Acad. Sci. U. S .A.* **93**, 2629–2634.

CHAPTER 5

APPENDIX

5.1 FAST FOLDING OF THE HIV-1 AND SIV GP41 SIX-HELIX BUNDLES

Daniel N Marti and Saša Bjelić and Min Lu and Hans Rudolf Bosshard and Ilian Jelesarov

Article published in J. Mol. Biol. 336: 1 - 8

To the following article I contributed data on the equilibrium stability and folding/unfolding kinetics of the SIV T586 variant.

doi:10.1016/j.jmb.2003.11.058 *J. Mol. Biol.* (2004) **336**, 1–8

JMB
Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER

COMMUNICATION

Fast Folding of the HIV-1 and SIV gp41 Six-helix Bundles

Daniel N. Marti[1], Saša Bjelić[1], Min Lu[2], Hans Rudolf Bosshard[1] and Ilian Jelesarov[1]*

[1]*Biochemisches Institut der*
Universität Zürich
Winterthurerstrasse 190
CH-8057 Zurich, Switzerland

[2]*Department of Biochemistry*
Weill Medical College of
Cornell University, New York
NY 10021, USA

**Corresponding author*

Human (HIV-1) and simian (SIV) immunodeficiency virus fusion with the host cell is promoted by the receptor-triggered refolding of the gp41 envelope protein into a stable trimer-of-hairpins structure that brings viral and cellular membranes into close proximity. The core of this hairpin structure is a six-helix bundle in which an inner homotrimeric coiled coil is buttressed by three antiparallel outer HR2 helices. We have used stopped-flow circular dichroism spectroscopy to characterize the unfolding and refolding kinetics of the six-helix bundle using the HIV-1 and SIV N34(L6)C28 polypeptides. In each case, the time-course of ellipticity changes in refolding experiments is well described by a simple two-state model involving the native trimer and the unfolded monomers. The unfolding free energy of the HIV-1 and SIV trimers and their urea dependence calculated from kinetic data are in very good agreement with data measured directly by isothermal unfolding experiments. Thus, formation of the gp41 six-helix bundle structure involves no detectable population of stable, partly folded intermediates. Folding of HIV-1 N34(L6)C28 is five orders of magnitudes faster than folding of its SIV counterpart in aqueous buffer: $k_{\mathrm{on,HIV\text{-}1}} = 1.3 \times 10^{15}\ \mathrm{M}^{-2}\ \mathrm{s}^{-1}$ *versus* $k_{\mathrm{on,SIV}} = 1.1 \times 10^{10}\ \mathrm{M}^{-2}\ \mathrm{s}^{-1}$. The unfolding rates are similar: $k_{\mathrm{off,HIV\text{-}1}} = 1.1 \times 10^{-5}\ \mathrm{s}^{-1}$ *versus* $k_{\mathrm{off,SIV}} = 5.7 \times 10^{-4}\ \mathrm{s}^{-1}$. Kinetic *m*-values indicate that the transition state for folding of the HIV-1 protein is significantly more compact than the transition state of the SIV protein. Replacement of a single SIV threonine by isoleucine corresponding to position 573 in the HIV-1 sequence significantly stabilizes the protein and renders the folding rate close to that of the HIV-1 protein yet without making the transition state of the mutant as compact as that of the HIV-1 protein. Therefore, the overall reduction of surface exposure in the high-energy transition state seems not to account for different folding rates. While the available biological evidence suggests that refolding of the gp41 protein is slow, our study implies that structural elements outside the trimer-of-hairpins limit the rate of HIV-1 fusion kinetics.

Keywords: folding kinetics; HIV-1 fusion; gp41; six-helix bundle; trimer-of-hairpins motif

---

Abbreviations used: HIV-1, human immunodeficiency virus type 1; HIV-1 protein, recombinant trimer-of-hairpins N34(L6)C28 construct derived from HIV-1 gp41 protein; SIV, simian immunodeficiency virus; SIV protein, recombinant trimer-of-hairpins N34(L6)C28 construct derived from SIV gp41 protein; HR, heptad-repeat region; CD, circular dichroism; K_{U}, unfolding equilibrium constant; ΔG_{U}, unfolding free energy; $\Delta G_{\mathrm{U}}^{\mathrm{H2O}}$, unfolding free energy in aqueous buffer; *m*, the first derivative of the observed unfolding free energy with respect to the denaturant concentration, units of kJ mol^{-1} M^{-1}; m_{on}, the first derivative of the logarithm of the folding rate constant with respect to the denaturant concentration, units of M^{-1}; m_{off}, the first derivative of the logarithm of the unfolding rate constant in respect to the denaturant concentration, units of M^{-1}.

E-mail address of the corresponding author: iljel@bioc.unizh.ch

In both human (HIV-1) and simian (SIV) immunodeficiency viruses, the envelope glycoprotein consists of two non-covalently associated subunits, gp120 and gp41, that are generated by proteolytic cleavage of a precursor protein, gp160.[1–3] Accumulating biochemical and structural evidence indicates that binding of gp120 to the CD4 cell-surface receptor triggers a series of conformational changes in the gp120/gp41 complex that promote recognition of the chemokine receptors and ultimately lead to fusion of the viral and cellular membranes.[4] The ectodomain of gp41 contains two heptad-repeat sequences (HR1 and HR2), one adjacent to the N-terminal fusion peptide and one near the C-terminal transmembrane segment. Peptides derived from the HR1 and HR2 regions associate to form a highly stable six-helix bundle in which three N-terminal HR1 helices form a central trimeric coiled coil, whereas three C-terminal HR2 helices pack in the antiparallel manner into three hydrophobic grooves on the surface of this coiled coil.[5–8] This six-helix bundle represents the core of the fusogenic gp41 trimer-of-hairpins structure (Figure 1). The gp41 protein is postulated to exist in at least three different conformations: (i) the native, metastable prefusogenic state, which is stabilized by extensive interactions with the gp120 surface subunit;[9–11] (ii) the transiently populated pre-hairpin intermediate, formed by exposure of the fusion peptide region and concurrent formation of the HR1 trimeric coiled coil;[12–15] and (iii) the fusogenic trimer-of-hairpins structure, in which the HR2 helices are associated with the HR1 coiled coil to appose the viral and cellular membranes for fusion.[5–8] While formation of the gp41 hairpin structure is an important thermodynamic driving force for the conformational activation of the envelope glycoprotein,[15] kinetic studies have shown that the resolution of the pre-hairpin intermediate to the trimer-of-hairpins form is slow.[12–14,16]

Here, we characterize the kinetics of gp41 six-helix bundle formation using the HIV-1 and SIV N34(L6)C28 recombinant models, which share the same topology but differ in sequence and have very different thermodynamic stabilities. Refolding kinetics and urea-induced equilibrium unfolding are consistently described by a simple two-state reaction involving the unfolded monomer and the native trimer. Detailed knowledge of the association kinetics of the six-helix bundle can provide insights into the folding mechanism of oligomeric proteins, in which association of the monomers and folding of the oligomer are intimately coupled. Furthermore, kinetic data may give clues about the dynamics and mechanistic steps that mediate formation of the fusion-active conformation. Our results show that the six-helix bundle formation is

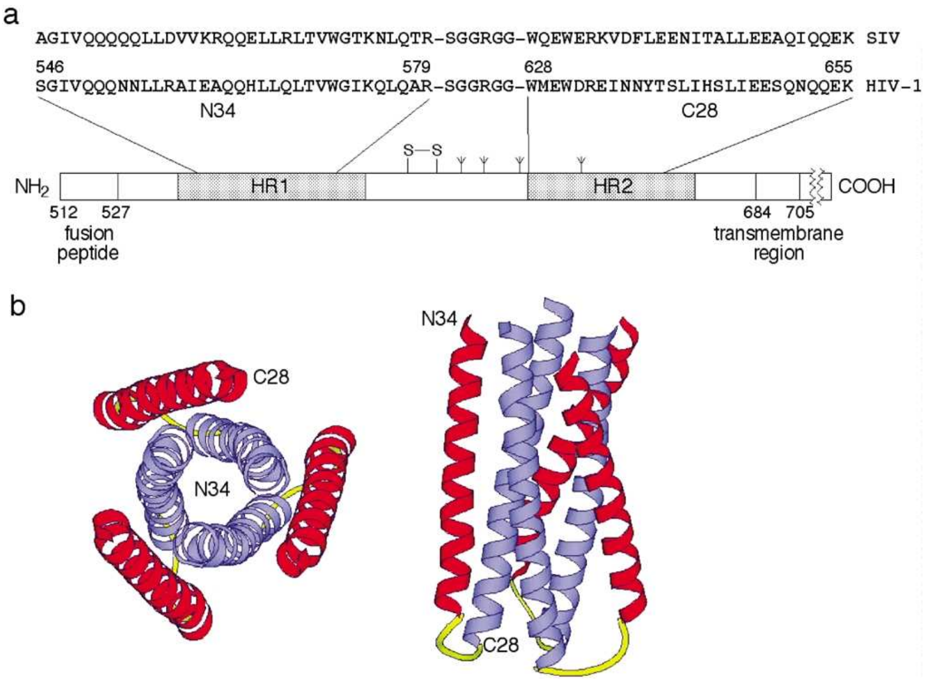


Figure 1. Core structure of HIV-1 gp41. a, Schematic view of the HIV-1 gp41 ectodomain. Two hydrophobic heptad-repeat sequences (HR1 and HR2) are indicated. The recombinant HIV-1 HXB2 N34(L6)C28 model consists of the N34 and C28 peptides connected *via* a six-residue linker that replaces the disulfide-bonded loop region of gp41. The sequence of the SIV N34(L6)C28 model is shown. The residues are numbered according to their position in gp160 of the HXB2 HIV-1 strain. b, Crystal structure of a six-helix bundle formed by HIV-1 N34(L6)C28. The left panel shows an axial view looking down the 3-fold axis of the six-helix bundle. The right panel shows a lateral view of the six-helix bundle.

not the rate-limiting step in the trimerization of HIV-1 and SIV gp41 and point to the presence of slower steps that mediate the *in vivo* fusion process.

Denaturant equilibrium unfolding of HIV-1 N34(L6)C28

The thermodynamic stability of the HIV-1 N34(L6)C28 trimer was investigated at 22 °C by urea denaturation in PBS, pH 7.0. (To facilitate reading we use protein or trimer instead of N34(L6)C28 throughout.) The midpoint of the urea-induced unfolding transition is 5.83 M for a total monomer concentration of 10.8 μM, as compared to a $[\text{urea}]_{1/2}$ of 1.48 M for SIV N34(L6)C28 at the same protein concentration (Figure 2). The experimental data are adequately described by a trimer-to-monomer two-state transition (Figure 2), as evaluated by non-linear optimization.[17] Our previous studies showed that the urea denaturation of the SIV protein and its Thr569-to-Ile and Thr573-to-Ile mutants (HIV-1 sequence numbering) follows a two-state mechanism.[17] The calculated unfolding free energy, ΔG_U^{H2O}, of the HIV-1 protein is 116(±10) kJ mol^{-1} (Table 1). Hence, the stability of the trimeric structure of the HIV-1 gp41 core is markedly increased relative to that of SIV trimer, for which a ΔG_U^{H2O}, of 79(±8) kJ mol^{-1} has been determined (Table 1,[17]). In the hydrophobic core formed by *a* and *d* position residues of the HR1 coiled coil trimer there is only a single non-conservative sequence difference between the SIV and HIV-1 protein: Ile573 of the HIV-1 protein is Thr573 in the SIV protein. Interestingly, the mutation Thr573Ile renders the SIV protein almost as stable as the HIV-1 protein (Table 1). The dependence of ΔG_U on the concentration of urea defined by m_{eq} remains as low as for the HIV-1 protein, ~10 kJ mol^{-1} M^{-1}. This is significantly lower m_{eq} of ~15 kJ mol^{-1} M^{-1} measured for the wild-type SIV protein (Table 1). Thus, the Thr573Ile mutation adjusts the hydrophobic core of the SIV protein (defined by heptad positions *a* and *d* in HR1) to that of the HIV-1 protein and changes the extent of surface exposure accompanying denaturation, for which m_{eq} is an indicator.[18] It seems that the unfolded state of both HIV-1 and mutant SIV protein is less extended than for the wild-type SIV protein.

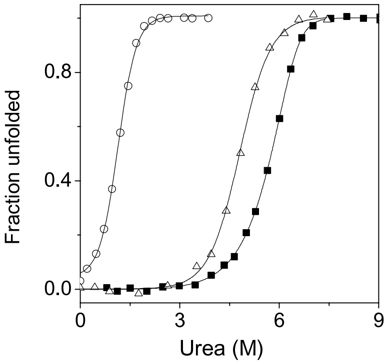


Figure 2. Urea-induced equilibrium unfolding transition of HIV-1 N34(L6)C28 (filled squares), SIV N34(L6)C28 (open circles) and SIV T573I N34(L6)C28 (open triangles) measured by the change in ellipticity at 222 nm at 22 °C. The continuous lines were computed by non-linear least squares fitting.[17] Total protein concentration was 10.7 μM in PBS (pH 7.0) and the corresponding urea concentration. CD measurements were performed on a Jasco-715 spectropolarimeter. Optical path length was 0.1 cm. The proteins were expressed in *Escherichia coli* strain BL21(DE3)/pLysS (Novagen) at 37 °C for three to four hours. Cells were lyzed by ice-cold glacial acetic acid. The bacterial lyzate was centrifuged (35,000*g* for 30 minutes) to separate the soluble fraction from inclusion bodies. The soluble fraction, containing the recombinant protein, was dialyzed against 5% (v/v) acetic acid for 12–16 hours at 4 °C. The protein was purified to homogeneity by reverse-phase high-performance liquid chromatography (HPLC) with a Vydac C_{18} preparative column and a linear water–acetonitrile gradient containing 0.1% (v/v) trifluoroacetic acid.

Refolding kinetics

Refolding rates of the HIV-1 and SIV proteins were determined by observing the change of the CD signal, $[\theta]_{225}$, after rapid dilution of the concentrated denaturant. The final urea concentrations after mixing were chosen to cover the range within which unfolding of the protein occurs, as established in the equilibrium unfolding experiments. Thus, CD refolding were collected between 1.48 M and 5.75 M urea for the HIV-1 protein, and between 0.28 M and 1.97 M urea for the SIV protein. The folding data were analyzed assuming a ternary folding mechanism:

$$3\text{M} \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} \text{T}$$

From a statistical point of view, simultaneous association of the three chains is very unlikely. It is more likely that a dimeric intermediate is formed before the third peptide chain associates according to the following coupled equilibria:

$$\text{M} + \text{M} \underset{k_{\text{off,D}}}{\overset{k_{\text{on,D}}}{\rightleftharpoons}} \text{D} + \text{M} \underset{k_{\text{off,T}}}{\overset{k_{\text{on,T}}}{\rightleftharpoons}} \text{T}$$

where the equilibrium unfolding constants of dimer and trimer decomposition are defined as:

$$K_{\text{U,D}} = \frac{[\text{M}]^2}{[\text{D}]} = \frac{k_{\text{off,D}}}{k_{\text{on,D}}}$$
$$\text{and } K_{\text{U,T}} = \frac{[\text{M}][\text{D}]}{[\text{T}]} = \frac{k_{\text{off,T}}}{k_{\text{on,T}}} \qquad (1a)$$

However, if dimer formation is not detected spectroscopically, trimer formation appears to be a ternary reaction with an overall equilibrium

Table 1. Equilibrium and kinetic parameters describing folding of HIV-1, SIV and SIV mutant proteins

Protein	Equilibrium parameters				Kinetic parameters				
	$\Delta G_{u}^{H2O a}$	${m_{eq}}^{a,b}$	$k_{on}^{H2O c}$	$k_{off}^{H2O c}$	$\Delta G_{u}^{H2O d}$	${m_{on}}^{c}$	${m_{off}}^{c}$	${m_{kin}}^{e}$	${\beta_T}^{f}$
HIV-1	116 ± 10	10.1 ± 1.0	$(1.31 \pm 0.25) \times 10^{15}$	$(1.08 \pm 0.54) \pm 10^{-5}$	113.5 ± 10.0	-3.45 ± 0.06	0.81 ± 0.09	10.5 ± 0.1	0.81
SIV	79 ± 8	15.0 ± 1.0	$(1.14 \pm 0.04) \times 10^{10}$	$(5.68 \pm 0.54) \times 10^{-4}$	75.5 ± 10.0	-3.57 ± 0.06	2.37 ± 0.09	14.6 ± 0.1	0.60
SIV T573I	110 ± 8	9.7 ± 1.0	$(1.08 \pm 0.08) \times 10^{14}$	$(6.98 \pm 0.49) \times 10^{-6}$	108 ± 10.0	-2.63 ± 0.06	1.39 ± 0.09	9.9 ± 0.1	0.65

At 22 °C in PBS (pH 7). ΔG_U in units of kJ mol^{-1}; m_{eq} and m_{kin} in units of kJ mol^{-1} M^{-1}; m_{on} and m_{off} in units of M^{-1}; k_{on} in units of M^{-2} s^{-1}; k_{off} in units of s^{-1}.

[a] Parameters obtained by non-linear optimization of a two-state model to the data.[17]

[b] Values for SIV are from Jelesarov & Lu.[17]

[c] Parameters obtained by linear fitting of equations (3a) and (3b).

[d] Calculated according to $\Delta Gu = -RT \ln(k_{off}/k_{on})$.

[e] Calculated according to $m = RT\ (|m_{kin}| + |m_{off}|)$ at 22 °C.

[f] β_T calculated according to $\beta_T = 1 - m_{off}/(|m_{on}| + |m_{off}|)$.

constant:

$$K_U = K_{U,T}K_{U,D} = \frac{[M]^3}{[T]} = \frac{k_{off,T}k_{off,D}}{k_{on,T}k_{on,D}} = \frac{k_{off}}{k_{on}} \quad (1b)$$

where $k_{on} = k_{on,T}k_{on,D}/k_{off,D}$ and $k_{off} = k_{off,T}$.

The refolding traces at a particular final urea concentration were fitted by numerical integration of the following equations:

$$\frac{d[M]}{dt} = -3k_{on}[M]^3 + 3k_{off}[T] \quad (2a)$$

$$\frac{d[T]}{dt} = k_{on}[M]^3 - k_{off}[T] \quad (2b)$$

The algorithm implemented in the program DynaFit was used.[19] The values of the refolding and unfolding rate constants k_{on} and k_{off}, were obtained simultaneously from fitting of the experimental signal. The kinetic traces were well reproduced when solving the differential equations derived for the ternary monomer-to-trimer reaction (Figure 3), indicating that trimerization of the HIV-1 and SIV proteins is indeed represented by a two-state reaction, in agreement with equilibrium unfolding data. If the proposed mechanism is correct, k_{on} and k_{off} should change exponentially with denaturant concentration according to:

$$\ln k_{on} = \ln k_{on}^{H2O} + m_{on}[\text{denaturant}] \quad (3a)$$

$$\ln k_{off} = \ln k_{off}^{H2O} + m_{off}[\text{denaturant}] \quad (3b)$$

This is indeed the case, as shown in Figure 4. The linear dependence of the rate constants holds in the urea concentration range 0.28–1.31 M for the SIV protein, and 3.26–5.75 M for the HIV-1 protein. At low urea concentrations, the folding reaction is too fast to be followed by stopped-flow. Likewise, the amplitude of the CD signal is too small at high urea concentration to accurately fit the folding curves: after reaching equilibrium, less than 45% of the monomeric chain is folded at the high limit urea concentrations. Assuming linear dependence on the denaturant concentrations, rate constants were extrapolated to zero denaturant to obtain and k_{off}^{H2O} (Table 1). Values of k_{off}^{H2O} differ by only one order of magnitude: $5.7 \times 10^{-4}\ s^{-1}$ for the SIV protein and $1.1 \times 10^{-5}\ s^{-1}$ for the HIV-1 protein. However, the folding rate constants are very different: $1.1 \times 10^{10}\ M^{-2}\ s^{-1}$ for SIV trimerization and $1.3 \times 10^{15}\ M^{-2}\ s^{-1}$ for HIV-1 trimerization. The equilibrium unfolding constant computed from kinetic data as $K_U^{H2O} = k_{off}^{H2O}/k_{on}^{H2O}$ is in excellent agreement with K_U^{H2O} derived from equilibrium unfolding data. Consequently, the free energies of unfolding ΔG_U^{H2O} derived from kinetic and equilibrium data also coincide very well (Table 1).

From the linear dependence of the kinetic rate constants on the urea concentration, m_{on} and m_{off} values were calculated (slopes of the regression lines in Figure 4). The m-values of the kinetic data contain information about the compactness of the

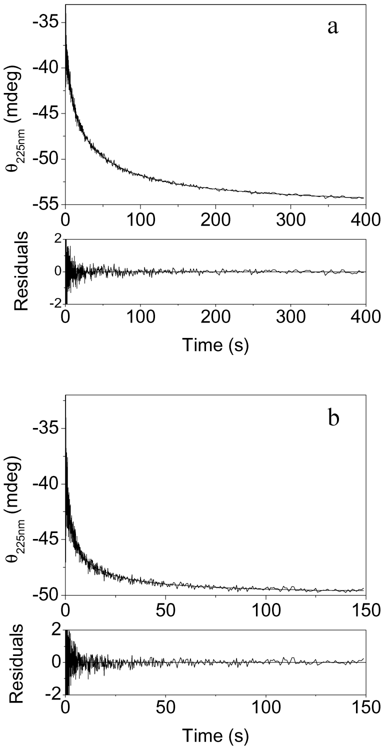


Figure 3. Refolding of the six-helix bundle proteins measured by CD stopped-flow. a, Representative folding trace of HIV-1 N34(L6)C28 (8.1 μM) in PBS (pH 7) plus 4.52 M urea. 89.1% of the molecule is folded after reaching equilibrium. b, Representative folding trace of SIV N34(L6)C28 (8.3 μM) in PBS (pH 7) plus 0.55 M urea. 89.0% of the molecule is folded after reaching equilibrium. The thick lines are best fits according to ternary reaction by numerical integration of equations (2a) and (2b). The residuals of the fits are shown in the lower panels. Fitted rate constants are shown in Figure 4. Note that the experiments (in a and b) were performed at different final urea concentrations and the apparent half-lives are not representative for the real kinetic differences between the proteins. Experiments were performed on an Applied Photophysics π*-180 stopped-flow spectrometer (dead time 1–2 ms). Refolding reaction was monitored by measuring the time dependence of the ellipticity at 225 nm at 22 °C using a path length of 10 mm and slit width of 4 nm. Typically, the results of two to four syringe firings were averaged for each kinetic trace. Protein stock solutions were prepared by dissolving HIV-1 N34(L6)C28 (88.7 μM) and SIV N34(L6)C28 (91.1 μM) in PBS (pH 7.0) plus 8 and 3 M urea, respectively. Refolding was initiated by a rapid tenfold dilution of the protein stock solution with PBS (pH 7.0) at a flow rate of 5 ml s^{-1}. Experiments were done at urea concentrations in the range 1.48–5.75 M and 0.28–1.97 M for HIV-1 N34(L6)C34 and SIV N34(L6)C28, respectively. The CD signal of the denatured protein at $t = 0$ was estimated by mixing the protein stock solutions with PBS (pH 7.0) containing 8 and 3 M urea, respectively.

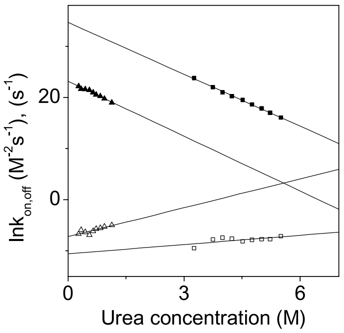


Figure 4. Urea dependence of the rate constants for refolding (filled symbols) and unfolding (open symbols) of HIV-1 N34(L6)C34 (squares) and SIV N34(L6)C28 (triangles). $\ln k_{\text{on}}$ and $\ln k_{\text{off}}$ were fitted from stopped-flow CD traces according to equations (2a) and (2b). Continuous lines were obtained by linear fitting of equations (3a) and (3b). The error associated with most of the data points is comparable with the size of the symbols. Fitted parameters are listed in Table 1.

transition state during folding, to be discussed below. For two-state folding, m-values from kinetics and equilibrium unfolding are related by $m_{\text{kin}} = RT(|m_{\text{on}}| + |m_{\text{off}}|)$. Indeed, kinetic and equilibrium m-values agree well (Table 1), indicating that six-helix bundle formation by HIV-1 and SIV proteins is well approximated by a ternary reaction mechanism and a two-state folding model.

The position of the transition state

In two-state folding, the reaction pathway between the unfolded and folded states leads through a high-energy conformation called transition state. The slopes m_{on} and m_{off} of the plots of $\ln k_{\text{on}}$ and $\ln k_{\text{off}}$ *versus* the denaturant concentration (Figure 4) are related to the degree of solvent exposure of residues of the transition state conformation relative to the folded and unfolded states.[18,20,21] The position of the transition state on the refolding reaction coordinate can be estimated from:

$$\beta_T = 1 - \frac{|m_{\text{off}}|}{|m_{\text{on}}| + |m_{\text{off}}|} \tag{4}$$

The β_T value is a relative measure of the solvent-exposed surface of the transition state relative to the folded state. A value of 1 would indicate that all of the residues buried in the folded state are already buried in the transition state. In this case, the transition state would be fully native-like and thus very compact relative to the unfolded state. The m-values derived from kinetic data suggest that the transition state is positioned towards the folded state on the reaction coordinates for both proteins. For the HIV-1 protein, ~80% of the surface buried in the folded trimer is already shielded in the transition state. The corresponding value for the SIV protein is ~60%. Hence, six-helix bundle folding, which is tightly coupled to association of monomers, involves traversing of a compact high-energy state, as it has been found for the majority of monomeric two-state folders.[21]

The role of the central coiled coil in determining the folding rate of the six-helix bundle

The five orders of magnitude difference in the folding rates of the HIV-1 and SIV six-helix bundles is intriguing. The overall structure of the two proteins is very similar, with virtually identical backbone topology. However, the HIV-1 protein is much more stable than the SIV counterpart, indicating pronounced energetic differences due to sequence differences that have only very minor effects on the three-dimensional structure. Replacement of Thr573 of the SIV protein by Ile strongly stabilizes the protein without changing the equilibrium two-state character of unfolding.[17] Thr573Ile is the only non-conservative difference between the HIV-1 and SIV proteins among nine *a* and *d* heptad positions building the hydrophobic core of the N-terminal coiled coil (blue central trimer in Figure 1). We determined the kinetic consequences of this single non-conservative mutation. Folding is accelerated and unfolding is slowed down in comparison to wild-type SIV. However, the folding rate is affected much more; the folding rate constant, $k_{\text{on}}^{\text{H2O}} = 1.1 \times 10^{14}\ \text{M}^{-2}\ \text{s}^{-1}$ of the SIV Thr573Ile mutant approaches closely that of the HIV-1 protein. Despite the similar folding rate constants, the transition state of the mutant appears as compact as the transition state of the wild-type protein ($\beta_T = 0.65$), both being less compact than the transition state of the HIV-1 protein (compare β_T values in Table 1). The overall surface burial in the transition state ensemble does not seem to correlate with the rate of folding as one might anticipate from the gain of energetically favorable contacts. We note that the refolding rate constant of the SIV Thr573Ile protein is less sensitive to the denaturant concentration, indicating structural differences in transition states of similar overall compactness.

The kinetic stability of the central trimeric coiled coil near position 573, which is close to the flexible loop region, seems to be a very important folding determinant. Possibly, structural consolidation and reduced dynamics of the heptad with Ile573 help to overcome the entropic penalty for fixing the loop region. At the same time, tighter packing and reduced dynamics may help to properly expose side-chains forming the grooves into which the HR2 helices pack (red helices in Figure 1). Since all *a* and *d* positions of the HR1 helices except position 573 are strictly conserved, it appears that formation of the central coiled coil is an early event in the folding process. The kinetic and thermodynamic differences between SIV protein and SIV Thr573Ile mutant imply that oligomeric proteins of the same topology fold with rates correlating with stability, much as found for monomeric two-state folders.[22]

Links to virus biology

It is widely believed that the refolding of gp41 into the fusogenic conformation is slow because the topological constraints preventing transition from the prehairpin intermediate to the trimer-of-hairpins structure by symmetric association of the HR1 and HR2 regions are slowly resolved.[7,12,14,16] The onset of cell–cell fusion is preceded by a lag phase at 37 °C of 15–20 minutes.[13,14,16] The slow *in vivo* kinetics of gp41 activation may explain why the fusion inhibitor T20 is capable of blocking HIV-1 fusion even at nanomolar concentrations, and even when added 15 minutes after the CD4 and co-receptor binding-triggered fusion events.[12,14,23,24] Significantly, a pre-hairpin intermediate is sensitive to T20 and to fusion-blocking antibodies, indicating that the HR1 coiled coil is formed and accessible at suboptimal temperatures, at which fusion does not occur or is retarded.[16,25] Thus, the pre-hairpin intermediate is induced rapidly upon CD4 binding and is then relatively stable.[12,14,16,26] Membrane fusion occurs concomitantly with six-helix bundle formation or shortly thereafter.[15] Altogether, HIV-1 fusion appears to be a relatively slow, stochastic process. A lag period prior to membrane fusion is thought to reflect the time needed to accumulate sufficient preformed six-helix bundles to form a fusion pore.

The present study demonstrates that there are no intrinsic barriers to rapid folding of the gp41 six-helix bundle. The results suggest that the kinetic barriers during HIV-1 and SIV fusion lie in the branch of the molecular pathway involving the pre-hairpin intermediate. The implication is that the HR1 and HR2 sequences *per se* are not rate-limiting in formation of the fusogenic trimer-of-hairpins. Instead, we suppose that the gp41 pre-hairpin intermediate structure must be unable to permit free diffusion of the complementary HR1 and HR2 regions. This could reflect the fact that the pre-hairpin intermediate spans both membranes, with the gp41 transmembrane domain anchored in the viral membrane and the fusion peptide inserted into the target cell membrane. Thus HR1 and HR2 are orders of magnitude slower to diffuse. Since the six-helix bundle model we are using does not include all of the structural elements within the gp41 ectodomain, it is possible that the additional sequence, for example, the membrane-proximal tryptophan-rich domain adjoining the HR2 region, could account for the remarkably slow HIV-1 fusion reaction.

Although the cascade of events leading to membrane fusion and virus entry cannot be comprehended from *in vitro* folding experiments, our experiments seem to capture some relevant properties of the six-helix bundle. Folding proceeds through a high-energy state, in which formation of the central coiled coil is likely advanced. Stability and overall surface burial achieved in that state are tuned by sequence differences. It has been demonstrated that point mutations in the HR1 helices correlate with the severity of *in vivo* phenotypes and with the variability of primary strains to T20 inhibition.[27,28] The simplicity of the model provides advantages in studying the energetic and kinetic principles of gp41 six-helix bundle formation.

Acknowledgements

This research was funded in part from grants from the Swiss National Science Foundation (to H.R.B. and I.J.) and NIH (grant AI42382 to M.L.).

References

1. Allan, J. S., Coligan, J. E., Barin, F., McLane, M. F., Sodroski, J. G., Rosen, C. A. *et al.* (1985). Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science*, **228**, 1091–1094.
2. Robey, W. G., Safai, B., Oroszlan, S., Arthur, L. O., Gonda, M. A., Gallo, R. C. & Fischinger, P. J. (1985). Characterization of envelope and core structural gene products of HTLV-III with sera from AIDS patients. *Science*, **228**, 593–595.
3. Burton, D. R. (2002). Antibodies, viruses and vaccines. *Nature Rev. Immunol.* **2**, 706–713.
4. Eckert, D. M. & Kim, P. S. (2001). Mechanisms of viral membrane fusion and its inhibition. *Ann. Rev. Biochem.* **70**, 777–810.
5. Lu, M., Blacklow, S. C. & Kim, P. S. (1995). A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nature Struct. Biol.* **2**, 1075–1082.
6. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell*, **89**, 263–273.
7. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J. & Wiley, D. C. (1997). Atomic structure of the ectodomain from HIV-1 gp41 [comment]. *Nature*, **387**, 426–430.
8. Tan, K., Liu, J., Wang, J., Shen, S. & Lu, M. (1997). Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl Acad. Sci. USA*, **94**, 12303–12308.
9. Thali, M., Moore, J. P., Furman, C., Charles, M., Ho, D. D., Robinson, J. & Sodroski, J. (1993). Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.* **67**, 3978–3988.
10. Sattentau, Q. J., Moore, J. P., Vignaux, F., Traincard, F. & Poignard, P. (1993). Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J. Virol.* **67**, 7383–7393.
11. Trkola, A., Dragic, T., Arthos, J., Binley, J. M., Olson, W. C., Allaway, G. P. *et al.* (1996). CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature*, **384**, 184–187.
12. Furuta, R. A., Wild, C. T., Weng, Y. & Weiss, C. D. (1998). Capture of an early fusion-active confor-

mation of HIV-1 gp41. *Nature Struct. Biol.* **5**, 276–279 (erratum appears in *Nature Struct. Biol.* (1998) **5**, 612).

13. Jones, P. L., Korte, T. & Blumenthal, R. (1998). Conformational changes in cell surface HIV-1 envelope glycoproteins are triggered by cooperation between cell surface CD4 and co-receptors. *J. Biol. Chem.* **273**, 404–409.
14. Munoz-Barroso, I., Durell, S., Sakaguchi, K., Appella, E. & Blumenthal, R. (1998). Dilation of the human immunodeficiency virus-1 envelope glycoprotein fusion pore revealed by the inhibitory action of a synthetic peptide from gp41. *J. Cell Biol.* **140**, 315–323.
15. Follis, K. E., Larson, S. J., Lu, M. & Nunberg, J. H. (2002). Genetic evidence that interhelical packing interactions in the gp41 core are critical for transition of the human immunodeficiency virus type 1 envelope glycoprotein to the fusion-active state. *J. Virol.* **76**, 7356–7362.
16. Melikyan, G. B., Markosyan, R. M., Hemmati, H., Delmedico, M. K., Lambert, D. M. & Cohen, F. S. (2000). Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion [comment]. *J. Cell Biol.* **151**, 413–423.
17. Jelesarov, I. & Lu, M. (2001). Thermodynamics of trimer-of-hairpins formation by the SIV gp41 envelope protein. *J. Mol. Biol.* **307**, 637–656.
18. Tanford, C. (1968). Protein denaturation. *Advan. Protein Chem.* **23**, 121–282.
19. Kuzmic, P. (1996). Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **237**, 260–273.
20. Tanford, C. (1970). Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Advan. Protein Chem.* **24**, 1–95.
21. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.
22. Fersht, A. R. (1999). Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl Acad. Sci. USA*, **97**, 1525–1529.
23. Wild, C. T., Shugars, D. C., Greenwell, T. K., McDanal, C. B. & Matthews, T. J. (1994). Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc. Natl Acad. Sci. USA*, **91**, 9770–9774.
24. Reeves, J. D., Gallo, S. A., Ahmad, N., Miamidian, J. L., Harvey, P. E., Sharron, M. *et al.* (2002). Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics. *Proc. Natl Acad. Sci. USA*, **99**, 16249–16254.
25. Golding, H., Zeitzeva, M., de Rosny, E., King, L. R., Manischewitz, J., Sidoeov, I. *et al.* (2002). Dissection of human immunodeficiency virus type 1 with neutralizing antibodies to gp41 fusion intermediates. *J. Virol.* **76**, 6780–6790.
26. Gallo, S. A., Puri, A. & Blumenthal, R. (2001). HIV-1 gp41 six-helix bundle formation occurs rapidly after the engagement of gp120 by CXCR4 in the HIV-1 Env-mediated fusion process. *Biochemistry*, **40**, 12231–12236.
27. Dubay, J. W., Roberts, S. J., Brody, B. & Hunter, E. (1992). Mutations in the leucine zipper of the human immunodeficiency virus type 1 transmembrane glycoprotein affect fusion and infectivity. *J. Virol.* **66**, 4748–4756.
28. Lu, M., Ji, H. & Shen, S. (1999). Subdomain folding and biological activity of the core structure from human immunodeficiency virus type 1 gp41: implications for viral membrane fusion. *J. Virol.* **73**, 4433–4438.

Edited by I. B. Holland

(Received 29 August 2003; received in revised form 28 November 2003; accepted 30 November 2003)

5.2 ENERGETICS OF PEPTIDE RECOGNITION BY THE SECOND PDZ DOMAIN OF HUMAN PROTEIN TYROSINE PHOSPHATASE 1E

Stoyan Milev and Saša Bjelić and Oleg Georgiev and Ilian Jelesarov
Article published in Biochemistry 46: 1064 - 1078

To the following article I contributed molecular dynamics (MD) simulations in explicit water of the human PTP1E second PDZ domain, the RA-GEF2-derived unstructured pentadecapeptide, and complex between the two. The analysis of the trajectories in terms of atomic fluctuations and changes in solvent accessible surface was indispensable in describing the thermodynamic complexity of the binding process.

Energetics of Peptide Recognition by the Second PDZ Domain of Human Protein Tyrosine Phosphatase 1E[†]

Stoyan Milev,[‡] Saša Bjelić,[‡] Oleg Georgiev,[§] and Ilian Jelesarov*,[‡]

Biochemisches Institut der Universität Zürich and Institut für Molekularbiologie der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received September 8, 2006; Revised Manuscript Received October 31, 2006

ABSTRACT: Formation of protein−protein assemblies is essential in maintaining cell structure and function. Conservation of structural motifs and binding sites is the result of evolutionary pressure for solutions compatible with both molecular economy and regulation. PDZ domains are a typical example: A conserved fold governs specificity toward recognition of C-terminal protein sequences by small sequential and/or structural deviations within a canonical binding mode. The energetic principles underlying the strength and specificity of PDZ−protein interactions are practically unknown. We use the second PDZ domain (PDZ2) of the human protein tyrosine phosphatase (hPTP1E) as a model to study the energetics of peptide binding to a class I PDZ domain. Calorimetric experiments reveal the enthalpy, entropy, and heat capacity changes accompanying PDZ2 binding to the C-terminal pentadecapeptide derived from the guanine nucleotide exchange factor RA-GEF2. Association is driven by favorable enthalpy and entropy changes below 18 °C. Above that temperature the entropy change opposes complex formation. Structure-based predictions poorly reproduce the observed thermodynamic profile of the PDZ−peptide complex. On the basis of MD simulations and experimental findings by others we suggest that changes in the dynamics of the PDZ domain upon peptide binding make a large contribution to the observed thermodynamic parameters. Possible impacts of subtle, ligand-induced structural "stiffening" of PDZ domains are discussed. In our hands, the C-terminal segment of the tumor suppressor APC binds much less tightly to PDZ2 than what has been proposed earlier from surface plasmon resonance experiments.

PDZ[1] domains are a family of small, evolutionary well-represented protein binding modules. Usually they are arranged in tandems and facilitate formation of multiprotein networks involved in signaling, cytoskeletal organization, and subcellular transport (*1*). Next to being "passive" scaffolds, PDZ domains can also modulate the function of partner proteins such as ion channels and membrane receptors (refs *2* and *3* and references cited therein). Mutations in genes encoding PDZ-containing proteins are linked to human diseases (*4*−*7*).

Despite low overall sequential homology the known high-resolution structures of PDZ domains exhibit only little structural variation (C_α RMSD on the order of 1.5 Å). Six β-strands form a β-sandwich with nonparallel planes. Two short α-helices pack on the edges of the β-sandwich to complement the hydrophobic core. PDZ domains recognize internal sequences of partner proteins in some cases, yet the typical binding mode is recognition of C-terminal segments. Binding takes place in a groove between strand β2 and helix α2. The peptide is fixed in an extended conformation, essentially complementing the β-sheet structure. The known PDZ/peptide structures show a conserved, canonical binding motif. The main anchoring point is the very C-terminal residue, which is invariantly hydrophobic. The apolar side chain fills the hydrophobic pocket formed between strands β2 and β3 and α-helix α2. The terminal carboxylate is coordinated by amide hydrogen bonds within a unique glycine-rich loop. Four upstream residues of the peptide ligand are typically involved in interactions with the protein. (Conventionally, sequence position 0 is assigned to the C-terminal residue of the peptide ligand and the sequence numbering is negative toward the N-terminus.) Depending on the specificity of interactions made by the residue located in position −2 of the incoming peptide with the side chain in position 1 of α-helix α2 of PDZ, three classes of PDZ domains have been defined so far (*3*). In class I PDZ, a hydrogen bond is formed between the hydroxyl group of Ser/Thr −2 and His 71. An apolar contact in this critical site defines class II PDZ. Aspartic acid or glutamic acid occupies position −2 in class III PDZ ligands. The preference is guided by the possibility of a hydrogen bond being formed between the side chain carboxylate and the hydroxyl group of tyrosine from helix α2. Examples violating this paradigm

---

[†] This work was supported in part by the Swiss National Science Foundation (Grant 31-100197/1).

* To whom correspondence should be addressed: phone, ++41 44 635 5547; fax, ++41 44 635 6805; e-mail, iljel@bioc.unizh.ch.

[‡] Biochemisches Institut der Universität Zürich.

[§] Institut für Molekularbiologie der Universität Zürich.

[1] Abbreviations: APC, tumor supressor adenomatous polyposis coli protein; ASA, solvent-accessible surface (in Å); ΔC_p, heat capacity change; DSC, differential scanning calorimetry; ΔG, free energy change; ΔH, enthalpy change; ITC, isothermal titration calorimetry; K_d, equilibrium dissociation constant; PDZ, postsynaptic density protein-95; PDZ2, the second PDZ domain of the human protein tyrosine phosphatase 1E (hPTP1E); RA-GEF2, guanine nucleotide exchange factor 2; ΔS, entropy change; RG, synthetic pentadecapeptide comprising the C-terminus of RA-GEF2.

10.1021/bi061869i CCC: $37.00

are documented, and some PDZ domains exhibit degenerate specificity (*8*). Positions −1 and −3 are more solvent exposed, exhibit no preference for specific side chains, and are thought to fine-tune the specificity and possibly the affinity of interactions. In some cases the PDZ domain provides anchoring sites for residues located as far as position −7 in the peptide ligand (*9*−*13*).

Although the structural determinants governing binding of C-terminal peptides to PDZ domains are now well established, much less is known about the energetics of association. Solid-phase methods such as surface plasmon resonance and ELISA yield dissociation constants typically in the range 1−200 nM (*14*−*16*). Differently, solution methods (ITC, fluorescence, NMR) suggest that the typical binding affinity may be lower, in the micromolar range (*17*−*20*), although a K_d of 270 nM was measured recently by ITC for a RIM1α PDZ domain with unusual specificity (*21*).

We select the second PDZ domain (PDZ2) of the human tyrosine phosphatase (hPTP1E) as a model to study the energetics of peptide binding to a class I PDZ domain. hPTP1E (alternatively known as PTPbas and PTPL1) is involved in maintaining the balance of tyrosine phosphorylation and is thus implicated in the regulation of diverse receptor-mediated signal transduction pathways, among them the regulation of cell growth and apoptosis in breast cancer (*22*, *23*). The protein contains five PDZ domains. The second domain, PDZ2, is known to interact with the human Fas/CD95 receptor (*24*), the zyxin-related protein ZRP-1 (*25*), the tumor supressor adenomatous polyposis coli protein (APC), and the guanine nucleotide exchange factor RA-GEF2 (*26*). C-Terminal peptides derived from the latter two proteins were reported to bind to PDZ with $K_d = 8$ nM [APC (*14*)] and $K_d = 10-30\ \mu$M [RA-GEF2 (*10*, *18*)], respectively.

The energetics of binding of C-terminal sequences derived from RA-GEF2 and APC proteins was characterized by isothermal titration calorimetry (ITC) and differential scanning calorimetry (DSC). The measurements confirm the low micromolar dissociation constant of the RA-GEF2-derived peptide (henceforth abbreviated as RG). Surprisingly, we found that the affinity of PDZ2 for the APC peptide is much lower than what has been suggested previously. To our knowledge, we report here for the first time the complete energetic profile of a PDZ−peptide complex. Analysis of the experimental thermodynamic parameters with respect to the predictions of semiempirical methods identifies energetic contributions from subtle structural changes in the PDZ domain induced by peptide binding.

EXPERIMENTAL PROCEDURES

Protein Cloning, Expression, and Purification. PDZ2 was PCR amplified from a human brain cDNA library by use of oligonucleotide primers designed from the DNA sequence (Genbank no. XM_172831). The primers PDZup (5′-GAATTcatatgCCTAAGCCTGGAGATATCTTTGAG-3′) and PDZdo-1 (5′-GCCggatccTCATGTTGGAGATTGTCCCTTTTCTAATAACAG-3′) contain *Nde*I and *Bam*HI sites, respectively. A stop codon was engineered into the PDZdo primer preceding the *Bam*HI site, so that a 96 amino acid fragment containing PDZ2 would be expressed. The PCR product was cloned between the *Nde*I and *Bam*HI sites of the pET21b bacterial expression vector and verified by sequencing. Overexpression of PDZ2 was achieved by growing *Escherichia coli* strain BL21(DE3) transformed with the expression vector in LB medium in the presence of ampicillin at 37 °C. Expression was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) ,and the cultures were grown for an additional 4 h. Cells were harvested by centrifugation, resuspended in 50 mM sodium acetate and 10 mM EDTA, pH 5.0, and stored at −80 °C. The resuspended cells were lyzed by sonification. The supernatant was loaded on a HiTrap SP HP cation-exchange column in buffer A (50 mM sodium acetate, 10 mM EDTA, pH 5.0), and the PDZ2 domain was eluted with buffer B (50 mM sodium acetate, 10 mM EDTA, 1 M NaCl, pH 5.0). The protein was further purified by reversed-phase HPLC in binary acetonitrile/water gradients containing 0.1% and 0.085% trifluoroacetic acid on a C8 column and lyophilized. The protein which was refolded from the lyophilized state was fully native as judged by spectroscopy and scanning calorimetry. The binding parameters obtained with material after cation exchange or HPLC as the final purification step were identical. The mass was verified by ESI mass spectrometry. Concentration was calculated using $\epsilon_{280} = 2125\ \mathrm{M}^{-1}\ \mathrm{cm}^{-1}$ determined by quantitative amino acid analysis.

Peptide Synthesis and Purification. Peptide Ac-NH-YADSEADENEQVSAV-OH (RG) corresponding to the 15 C-terminal amino acids of guanine nucleotide exchange factor RA-GEF-2 and peptide Ac-NH-SSGTQSPKRHSGSYLVTSV-OH (APC) corresponding to the 19 C-terminal amino acids of tumor suppressor APC were custom synthesized using the N^{α}-Fmoc protection strategy. The naturally occurring phenylalanine in RG (N-terminal residue in the pentadecapeptide) was replaced by tyrosine to facilitate concentration determination by UV. The N-terminal residue was acetylated by reaction of the resin-bound and side chain protected peptide with a 10-fold molar excess of acetic anhydride and a 5-fold molar excess of trimethylamine in dimethylformamide. After deprotection and cleavage from the resin, crude peptide preparations were desalted on a Sephadex G-25 column in 1 M acetic acid. Final purification was achieved by reversed-phase HPLC on a semipreparative C8 column eluted with binary acetonitrile/water gradients containing 0.1% trifluoroacetic acid. The purity of peptides was controlled by ESI mass spectrometry. The concentration was determined by UV absorption using $\epsilon_{280} = 1280\ \mathrm{M}^{-1}\ \mathrm{cm}^{-1}$.

Buffer. All experiments were conducted in standard buffer composed of 50 mM sodium phosphate and 150 mM NaCl, pH 6.8. The pH of samples containing urea was adjusted after addition of the denaturant. Urea concentrations were determined by measuring the refraction index. All chemicals were of analytical grade and were used without further purification.

Differential Scanning Calorimetry (DSC). DSC experiments were performed on a VP-DSC calorimeter (MicroCal Inc.) equipped with twin coin-shaped cells of 0.52 mL volume. Details on the instrument's performance are given elsewhere (*27*). The heating rate was 1 °C min^{-1}. Samples containing protein and peptide (in isolation or as a mixture) were dialyzed for 18−24 h against the same batch of buffer used to establish the instrumental buffer−buffer baseline. Reversibility was checked by two to three cycles of heating and cooling. The raw experimental data were corrected for the instrumental baseline and transformed to partial molar or partial specific heat capacity using partial specific volumes of 0.724, 0.684, and 0.718 cm^3 g^{-1} for PDZ2, RG, and the PDZ2/RG mixture, respectively, calculated from the amino acid composition. The analysis of heat capacity traces of the protein−peptide complex followed the formalism detailed elsewhere (*28*−*30*). Briefly, considering the unfolding of a monomeric protein, the temperature dependence of the measured heat capacity is expressed as

$$C_p(T) = C_{p,\mathrm{N}} + f_\mathrm{U}\Delta C_p + \Delta H_\mathrm{m}\frac{\mathrm{d}f_\mathrm{U}}{\mathrm{d}T} = C_{p,\mathrm{N}} + f_\mathrm{U}\Delta C_p + \frac{K_\mathrm{U}}{(1+K_\mathrm{U})^2}\left(\frac{\Delta H_\mathrm{m}{}^2}{RT^2}\right) \quad (1)$$

where K_U is the equilibrium unfolding constant, $f_\mathrm{U} = K_\mathrm{U}/(1 + K_\mathrm{U})$ is the fraction of unfolded protein, ΔH_m is the unfolding enthalpy at T_m, the temperature where $f_\mathrm{U} = 0.5$ (i.e., $K_\mathrm{U} = 1$), and $\Delta C_p = C_{p,\mathrm{U}} - C_{p,\mathrm{N}}$ is the unfolding heat capacity change. The heat capacity of the folded protein, $C_{p,\mathrm{N}}$, was modeled with a linear function. As shown before, the heat capacity of the unfolded state, $C_{p,\mathrm{U}}$, can be calculated with good precision from the amino acid composition of the protein and is well approximated by a second-order polynomial function of the general form $C_{p,\mathrm{U}} = a + bT + cT^2$, where a, b, and c are coefficients (*31*). In eq 1, it is implicitly assumed that K_U depends on the temperature according to eq 6 below. Regression analysis according to eq 1 returns an optimized value for ΔH_m, which is in fact the geometric mean of the model-independent calorimetric enthalpy, ΔH_cal, and the model-dependent van't Hoff enthalpy, ΔH_vH, i.e., $\Delta H_\mathrm{m} = (\Delta H_\mathrm{cal}\Delta H_\mathrm{vH})^{1/2}$. For a two-state transition, $\Delta H_\mathrm{cal} = \Delta H_\mathrm{vH}$. The calorimetric enthalpy is obtained by integration of the $C_p(T)$ function above the intrinsic heat capacity change function, which is defined by the first two terms on the right-hand side of eq 1. The van't Hoff enthalpy can be calculated as $\Delta H_\mathrm{vH} = 2T_\mathrm{m}[R(C_{p,\mathrm{max}} - \Delta C_p/2)]^{1/2}$, where $C_{p,\mathrm{max}}$ is the heat capacity at T_m. Data handling and analysis were carried out using the program CpCalc 2.1 (Applied Thermodynamics), subroutines for Origin provided by MicroCal, and in-house written scripts for NLREG (Phillip H. Sherrod).

Isothermal Titration Calorimetry (ITC). ITC experiments were performed on a VP-ITC instrument (MicroCal Inc.). The calorimeter was calibrated according to the manufacturer's instruction. Samples of protein and peptide were prepared in, and thoroughly dialyzed against, the same batch of buffer to minimize artifacts due to minor differences in buffer composition. The concentration was determined after dialysis. The sample cell (1.4 mL) was loaded with 30−80 μM protein; peptide concentration in the syringe was 400−1000 μM. A titration experiment typically consisted of 25−30 injections, each of 8 or 10 μL volume and 10 or 12 s duration, with a 5 min interval between additions. The stirring rate was 300 rpm. Raw data were integrated, corrected for nonspecific heats, normalized for concentration, and analyzed according to a 1:1 binding model assuming a single set of identical binding sites.

Circular Dichroism Spectroscopy (*CD*). CD measurements were carried out on a Jasco J-715 instrument equipped with a computer-controlled water bath using jacketed cuvettes of 0.1 and 1.0 cm optical path. Isothermal urea unfolding experiments were performed at 5, 15, 25, and 35 °C. The data were analyzed following the linear extrapolation model (LEM) as detailed elsewhere (*32*). The ΔG_U values measured at different temperatures define the stability curve of the protein according to the Gibbs−Helmholtz equation:

$$\Delta G_\mathrm{U}(T) = \Delta H_\mathrm{R}\left(1 - \frac{T}{T_\mathrm{R}}\right) + \Delta C_p\left[T - T_\mathrm{R} - T\ln\left(\frac{T}{T_\mathrm{R}}\right)\right] \quad (2)$$

Conveniently, T_R is selected as the temperature where the protein is half unfolded ($f_\mathrm{U} = 0.5$ and $\Delta H_\mathrm{U} = \Delta H_\mathrm{R}$). For monomeric proteins, if T_R is known from an independent experiment (thermal unfolding), $\Delta G(T_\mathrm{R}) = -RT\ln[f_\mathrm{U}/(1 - f_\mathrm{U})] = 0$ can be included in the data set to reduce the uncertainty when three parameters are being optimized based on (usually) few experimental points collected at low temperatures.

Binding Simulations. To simulate the temperature dependence of the excess heat capacity function of a mixture of protein and peptide, the theoretical framework developed by Brandts and Lin was used (*33*). We assume that a 1:1 complex between protein and peptide is formed; the peptide binds only to folded protein. The relevant equilibria to be considered are

$$\mathrm{PDZ_F} + \mathrm{RG} \overset{K_\mathrm{A}}{\rightleftharpoons} \mathrm{PDZ_F/RG}$$

$$\mathrm{PDZ_F} \overset{K_\mathrm{U}}{\rightleftharpoons} \mathrm{PDZ_U}$$

The equilibrium association constant, K_A, and the equilibrium unfolding constant are defined in the usual way as

$$K_\mathrm{A} = \frac{[\mathrm{PDZ_F/RG}]}{[\mathrm{PDZ_F}][\mathrm{RG}]} \quad (3)$$

$$K_\mathrm{U} = \frac{[\mathrm{PDZ_U}]}{[\mathrm{PDZ_F}]} \quad (4)$$

The brackets indicate the equilibrium concentrations of folded unbound protein ($\mathrm{PDZ_F}$), unfolded protein ($\mathrm{PDZ_U}$), unbound peptide (RG), and the complex ($\mathrm{PDZ_F/RG}$). The temperature dependencies of K_A and K_U are given by

$$K_\mathrm{A}(T) = K_\mathrm{A,R}\exp\left[-\frac{\Delta H_\mathrm{A,R}}{R}\left(\frac{1}{T} - \frac{1}{T_\mathrm{R}}\right) + \frac{\Delta C_{p,\mathrm{A}}}{R}\left(\ln\frac{T}{T_\mathrm{R}} + \frac{T_\mathrm{R}}{T} - 1\right)\right] \quad (5)$$

$$K_\mathrm{U}(T) = \exp\left[-\frac{\Delta H_\mathrm{m}}{RT}\left(1 - \frac{T}{T_\mathrm{m}}\right) - \frac{\Delta C_p}{RT}\left(T - T_\mathrm{m} - T\ln\frac{T}{T_\mathrm{m}}\right)\right] \quad (6)$$

$K_\mathrm{A,R}$ and $\Delta H_\mathrm{A,R}$ are the binding constant and the binding enthalpy change at the reference temperature T_R, which was taken as 298.15 K (25 °C), and $\Delta C_{p,\mathrm{A}} = \mathrm{d}\Delta H_\mathrm{A}/\mathrm{d}T$ is the heat capacity change of association. Parameters which characterize unfolding of the protein [ΔH_m, ΔC_p, and T_m = 322.6 K (49.4 °C)] are defined as in eq 2. At any temperature, the enthalpy and heat capacity changes can be calculated from

$$\Delta H_{\mathrm{A}} = \Delta H_{\mathrm{A,R}} + \Delta C_{p,\mathrm{A}}(T - T_{\mathrm{R}}) \qquad (7)$$

$$\Delta H_{\mathrm{U}} = \Delta H_{\mathrm{m}} + \Delta C_{p}(T - T_{\mathrm{m}}) \qquad (8)$$

The mass conservation equations link the two equilibria. For the total protein ($\mathrm{PDZ_T}$) and peptide ($\mathrm{RG_T}$) concentration we can write

$$[\mathrm{PDZ_T}] = [\mathrm{PDZ_F}] + [\mathrm{PDZ_U}] + [\mathrm{PDZ_F/RG}] = [\mathrm{PDZ_F}] + K_{\mathrm{U}}[\mathrm{PDZ_F}] + K_{\mathrm{A}}[\mathrm{PDZ_F}][\mathrm{RG}] \qquad (9)$$

$$[\mathrm{RG_T}] = [\mathrm{RG}] + [\mathrm{PDZ_F/GR}] = [\mathrm{RG}] + K_{\mathrm{A}}[\mathrm{PDZ_F}][\mathrm{RG}] \qquad (10)$$

Since $\mathrm{PDZ_T}$ and $\mathrm{RG_T}$ are known, the combined eqs 3−10 can be solved simultaneously to determine the concentrations of the relevant species at any temperature:

$$[\mathrm{PDZ_F}] = [\mathrm{PDZ_T}]/(1 + K_{\mathrm{U}} + K_{\mathrm{A}}[\mathrm{RG}]) \qquad (11)$$

$$[\mathrm{PDZ_U}] = K_{\mathrm{U}}[\mathrm{PDZ_T}]/(1 + K_{\mathrm{U}} + K_{\mathrm{A}}[\mathrm{RG}]) \qquad (12)$$

$$[\mathrm{PDZ_F/RG}] = K_{\mathrm{U}}[\mathrm{RG}][\mathrm{PDZ_T}]/(1 + K_{\mathrm{U}} + K_{\mathrm{A}}[\mathrm{RG}]) \qquad (13)$$

Taking the protein−peptide complex as the reference state, the excess enthalpy function, $\langle \Delta H \rangle$, is defined as

$$\langle \Delta H \rangle = \Delta H_{\mathrm{U}} \frac{[\mathrm{PDZ_U}]}{[\mathrm{PDZ_T}]} - \Delta H_{\mathrm{A}} \frac{[\mathrm{PDZ_F}] + [\mathrm{PDZ_U}]}{[\mathrm{PDZ_T}]} \qquad (14)$$

The minus sign accounts for complex dissociation. Numerical differentiation of eq 14 yields the temperature dependence of the excess heat capacity function.

Electrostatic Modeling. The free protein was modeled either with the structure extracted from the NMR ensemble of the complex or with the NMR ensembles of the free human PDZ2 (3PDZ) and the free mouse homologue (1GM1; 94% sequence identity). Ten conformers of the corresponding NMR ensembles were used. The calculation was done with the Poisson−Boltzmann model in the framework of the continuum approximation, as implemented in the program MEAD (*34*).

Molecular Dynamics Simulations. The MD simulations were carried out with the GROMACS simulation suite (version 3.3.1) (*35*) using the OPLS all-atom force field. Models 1 and 20 of the NMR ensemble (1D5G) were used as the starting point for MD simulation. The clashes in the structure were removed with the WHATIF program (http://swift.cmbi.kun.nl/WIWWWI/). The structure (complex, protein, or peptide) was solvated with TIP4 water (*36*), including approximately 150 mM NaCl (plus additional ions to neutralize the total system). A cubic periodic box was chosen such that the minimum distance between the protein and the end of the box was more the 15 Å. After minimization using the steepest descent algorithm with a tolerance of 100 kJ mol^{-1} Å^{-1}, the system solvent was relaxed for 500 ps with a harmonic position restraint on all C_α atoms (force constant: 10 kJ mol^{-1} Å^{-2}). LINCS (*37*) and SETTLE (*38*) algorithms were applied for constraining the bond lengths. The integration step was 2 fs. Short-range electrostatics were calculated explicitly, and long-range electrostatic interactions were calculated using the particle-mesh Ewald method (*39*). A 10 Å cutoff was used for van der Waals interactions. A long-range correction for the energy and the pressure was applied. The system was coupled to a Berendsen temperature bath ($\tau_t = 0.1$ ps), separately for the protein and the solvent, and to a Berendsen pressure bath ($\tau_p = 0.1$ ps) (*40*). The simulations were performed at 300 K. The simulation time was 15 ns. Trajectory visualization and analyzing was made with the visualization and analyzing software VMD.

Structural Parametrization. In the following we briefly describe the methods used for calculating of energetic terms based on the structure of the PDZ2/RG complex. Detailed discussion can be found elsewhere (*41*, *42*). Solvent-accessible surface (ASA) calculations were performed with the program NACCESS using a probe radius of 1.4 Å and a slice width of 0.25 Å, the default set of atomic radii (*43*). From each trajectory, the average ASA was calculated as the arithmetic mean over ASAs of equally distant snapshots (step 20 ps) taken in the last 10 ns of simulation. The ensemble-averaged differences in ASA between the associated and dissociated state (ΔASA) was partitioned into polar ($\Delta\mathrm{ASA_{pol}}$) and apolar ($\Delta\mathrm{ASA_{apol}}$) components. The hydration contribution to the binding heat capacity (ΔC_p) is calculated by

$$\Delta C_p = \sum C_{p,\mathrm{i}} \Delta\mathrm{ASA_i} \qquad (15)$$

where the terms ΔASA are in units of Å^2 and the coefficients $C_{p,\mathrm{i}}$ are the elementary contributions (in units of kJ K^{-1} mol^{-1} Å^{-2}) to the heat capacity of hydration of the corresponding type of surface. The following $C_{p,\mathrm{i}}$ values were used: $C_{p,\mathrm{apol}} = 1.82$, $C_{p,\mathrm{pol}} = -1.09$, and $C_{p,\mathrm{OH}} = 0.71$ (the last term accounts for the heat capacity contribution of buried hydroxyl groups of serine and threonine). In the absence of proton release/uptake upon complex formation, the generic binding enthalpy at 60 °C can be calculated within an estimated error range of 10% as (in kJ mol^{-1}):

$$\Delta H^{60°}_{\mathrm{calc}} = 122\Delta\mathrm{ASA_{pol,i}} - 30.4\Delta\mathrm{ASA_{apol}} \qquad (16)$$

At any other temperature the binding enthalpy is determined using $\Delta H^{60°}$ and the calculated ΔC_p.

It has naturalized itself to partition the binding entropy change into three terms describing (i) the entropy of water reorganization caused by the change in molecular surface (ΔS_{hyd}), (ii) the loss of conformational entropy due to side chain and backbone immobilization (ΔS_{conf}), and (iii) the change of rotational/translational degrees of freedom when two free kinetic units form the bimolecular complex (ΔS_{rt}):

$$\Delta S = \Delta S_{\mathrm{hyd}} + \Delta S_{\mathrm{conf}} + \Delta S_{\mathrm{rt}} \qquad (17)$$

The change in hydration entropy is calculated by extrapolation, using ΔC_p and the temperatures where the hydration entropy of polar and apolar atoms is zero (335.15 and 385.15 K, respectively):

$$\Delta S_{\mathrm{hyd}} = C_{p,\mathrm{apol}} \Delta\mathrm{ASA_{apol}} \ln\left(\frac{T}{385.15}\right) + C_{p,\mathrm{pol}} \Delta\mathrm{ASA_{pol}} \ln\left(\frac{T}{335.15}\right) \qquad (18)$$

The calculation of the ΔS_{conf} term of eq 17 poses a major problem in estimating energetics from structural data. According to the method used here, the side chain entropy loss can be estimated from the change in $\mathrm{ASA_i}$ upon binding:

$$\Delta S_{\mathrm{conf,SC}} = \sum_i \frac{\Delta \mathrm{ASA_i}}{\mathrm{ASA_{AXA,i}}} \Delta S_{\mathrm{bu-ex,i}} \qquad (19)$$

$\mathrm{ASA_{AXA,i}}$ is ASA of the corresponding side chain in the fully exposed state (modeled as Ala-X-Ala), and $\Delta S_{\mathrm{bu-ex,i}}$ is the entropy loss for a complete burial of the side chain of type i. Residues immobilized in the binding pocket experience loss of conformational entropy from backbone immobilization (ΔS_{bb}). An additional term ($\Delta S_{\mathrm{ex-u}}$) represents the unfavorable change in entropy for side chains which are solvent exposed but structured in the protein−peptide complex. The values of $\mathrm{ASA_{AXA,i}}$, $\Delta S_{\mathrm{bu-ex,i}}$, ΔS_{bb}, and $\Delta S_{\mathrm{ex-u}}$ were taken from ref *41*.

The magnitude of ΔS_{rt} is still uncertain (ref *44* and cited work therein). We use here $\Delta S_{\mathrm{rt}} = -35$ J K^{-1} mol^{-1}. This number is numerically close to the "cratic entropy" and has been found to describe reasonably the loss of rotational/translational degrees of freedom in diverse experimental systems (*44*−*46*).

RESULTS AND DISCUSSION

PDZ2 was expressed in soluble form in the cytoplasm of *E. coli*. The purified protein is monomeric at the concentrations and under the experimental conditions used in this study, as evidenced by light scattering experiments. The far-UV CD spectrum has the spectral signature of proteins containing predominantly β-sheets. CD spectroscopy demonstrates that the two peptides used in this study adopt a random coil conformation in solution (not shown). We first describe thermal and isothermal urea-induced unfolding experiments aimed at determination of the thermodynamic stability and clarification of the unfolding mechanism of PDZ2. This information is important in order to confine binding experiments in the temperature range where PDZ2 is fully native, as well as to interpret the results from thermal melting of the protein−peptide complex. ITC and DSC experiments revealing the energetics of peptide binding to the protein are presented in the second part. In the third part we search for correlations between the observed thermodynamic parameters and the structural features of the PDZ2/RG complex.

Stability and Unfolding Energetics of PDZ2. (A) Thermal Unfolding. The thermodynamic stability and unfolding mechanism of PDZ2 were characterized by DSC (Figure 1). The partial molar heat capacity at 25 °C is 15.6 ± 1.0 kJ K^{-1} mol^{-1}, corresponding to a partial specific heat capacity of 1.56 ± 1.00 J K^{-1} g^{-1}. This value is typical for globular proteins (*47*). Between 5 and 30 °C, the heat capacity increases linearly with a slope $\mathrm{d}C_p/\mathrm{d}T = 0.107 \pm 0.005$ kJ K^{-2} mol^{-1}, corresponding to $(10.7 \pm 0.5) \times 10^{-3}$ J K^{-2} g^{-1}. Although not extraordinarily steep in comparison with compact globular proteins (*47*), this increase indicates some degree of flexibility, possibly caused by thermal fluctuations of the disordered N- and C-termini (*10*). The heat capacity at 75 °C closely matches the predicted heat capacity based on summation of the heat capacities of the constituent amino acids (*31*). Hence, thermal unfolding results in a highly hydrated state lacking residual tertiary interactions. Integration of the heat absorption peak above the chemical baseline connecting the pretransitional and posttransitional heat capacities in proportion to the extent of unfolding yields the model-independent, calorimetric parameters characterizing PDZ2 unfolding: $T_{\mathrm{m,P}} = 49.4$ °C, $\Delta H_{\mathrm{cal,P}}(T_{\mathrm{m}}) = 273 \pm 10$ kJ mol^{-1}, $\Delta S_{\mathrm{cal,P}}(T_{\mathrm{m}}) = 846 \pm 30$ J K^{-1} mol^{-1}, and $\Delta C_{p,\mathrm{P}} = 4.45 \pm 0.40$ kJ K^{-1} mol^{-1} (Table 1). From a pH-induced variation of ΔH_{m} in a limited range of T_{m}, Gianni et al. estimated $\Delta C_p = 5.0 \pm 0.6$ kJ K^{-1} mol^{-1} for the PDZ2 Y43W mutant (*48*). The specific unfolding parameters at 60 °C are $\Delta H_{\mathrm{cal,P}}(60) = 3.3 \pm 2$ kJ mol res^{-1}, $\Delta S_{\mathrm{cal,P}}(60) = 10.3 \pm 0.5$ J K^{-1} mol res^{-1}, and $\Delta C_{p,\mathrm{P}} = 46 \pm 2$ J K^{-1} mol res^{-1} and are well within the range of values determined for globular proteins (*49*, *50*). Altogether, the calorimetric data imply that PDZ2 is a compact globular domain.

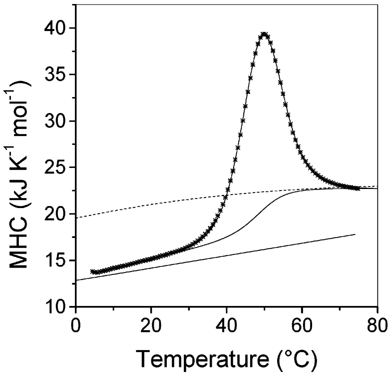


FIGURE 1: Thermal unfolding of PDZ2 observed by DSC. The experimentally measured temperature dependence of the partial molar heat capacity function is shown with symbols (average of three individual scans). The intrinsic heat capacity change (chemical baseline; continuous line below the heat absorption peak) was calculated as $C_{p,\mathrm{base}} = (1 - f_{\mathrm{U}})C_{p,\mathrm{N}} + f_{\mathrm{U}}C_{p,\mathrm{U}}$, where $C_{p,\mathrm{N}}$ and $C_{p,\mathrm{U}}$ are the heat capacities of the native state and of the unfolded state, respectively, and f_{U} is the fractional amount of heat absorbed between 30 and 75 °C. The heat capacity of the denatured state calculated by summation of the heat capacities of the constituent amino acids is indicated by the dotted line. The straight line represents the heat capacity of an average native globular protein with the mass of PDZ2. The line associated with the symbols visualizes the result of the nonlinear least-squares regression analysis. Statistically indistinguishable fits were obtained assuming a two-state transition ($\Delta H_{\mathrm{fit}} = 256 \pm 10$ kJ mol^{-1}) or non-two-state transition ($\Delta H_{\mathrm{cal}} = 273 \pm 10$ kJ mol^{-1}, $\Delta H_{\mathrm{vH}} = 240 \pm 10$ kJ mol^{-1}). ΔH_{fit} is identical within error with the weighted enthalpy estimate, $\Delta H_{\mathrm{WA}} = 0.65\Delta H_{\mathrm{vH}} + 0.35\Delta H_{\mathrm{cal}} = 251 \pm 14$ kJ mol^{-1}. Experiments were performed with 130−220 μM protein in 50 mM sodium phosphate and 150 mM NaCl, pH 6.8.

Analysis of the molar heat capacity function indicates an apparent small deviation from the two-state unfolding behavior. The temperature dependence of the excess heat capacity is best described by the model-dependent van't Hoff enthalpy $\Delta H_{\mathrm{vH}} = 240 \pm 10$ kJ mol^{-1}. Thus the ratio $\Delta H_{\mathrm{vH,DSC}}/\Delta H_{\mathrm{cal}}$ is 0.89 ± 0.06. However, thermodynamic modeling failed to detect the presence of statistically robust intermediates. We assume, therefore, that the apparent deviation from the two-state model is caused by accumulation of experimental errors. On the other hand, it has been proposed that a substantial improvement in the accuracy of ΔH at T_{m} can be achieved using a weighted enthalpy average, which is calculated as $\Delta H_{\mathrm{WA}} = 0.65\Delta H_{\mathrm{vH}} + 0.35\Delta H_{\mathrm{cal}}$ (*51*).

Table 1: Thermodynamic Parameters Determining the Stability of PDZ2[a]

T	ΔH	ΔS	ΔC_p	ΔG	$U_{1/2}$	$-m_{\mathrm{eq}}$
5[b]				22.1 ± 1.6	3.2 ± 0.1	6.9 ± 0.5
8[c]				21.7 ± 2.1	2.1[c]	10[c]
15[b]				19.2 ± 2.6	2.94 ± 0.14	6.52 ± 0.9
25[b]				14.6 ± 1.3	2.65 ± 0.12	5.52 ± 0. 5
35[b]				10.6 ± 1.3	2.11 ± 0.17	4.99 ± 0.6
49.4[d]	273 ± 10[e]	0.846 ± 0.030[e]	4.45 ± 0.40[e]	0 ± 14[e]		
	240 ± 10[f]					
	251 ± 14[g]					
	255 ± 32[h]		4.35 ± 1.90[h]			

[a] All experiments in standard buffer, 50 mM sodium phosphate and 150 mM NaCl, pH 6.8. ΔH and ΔG in units of kJ mol^{-1}; ΔS and ΔC_p in units of kJ K^{-1} mol^{-1}; $U_{1/2}$ in units of M; m_{eq} in units of kJ mol^{-1} M^{-1}. [b] From urea-induced unfolding measured by circular dichroism. [c] From GdmCl-induced unfolding measured by fluorescence (*17*). $U_{1/2}$ and m_{eq} are rough estimates taken from the corresponding plot in this paper. [d] From DSC. [e] Calorimetric estimate. [f] van't Hoff estimate. [g] Calculated as $0.65\Delta H_{\mathrm{vH}} + 0.35\Delta H_{\mathrm{cal}}$. [h] Calculated as the fitting parameter in the Gibbs−Helmholtz equation.

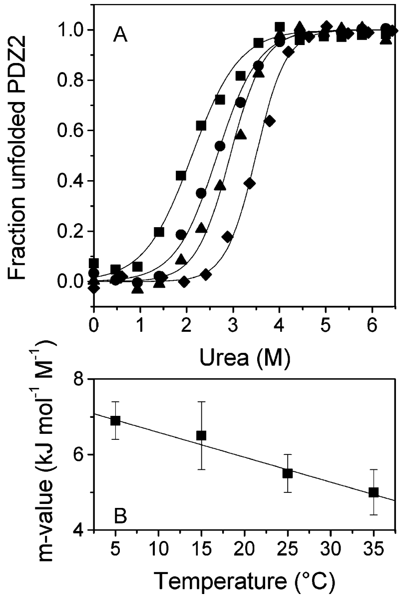


FIGURE 2: PDZ2 unfolding induced by urea-monitored CD spectroscopy. (A) Fraction of unfolded protein at 5 °C (diamonds), 15 °C (triangles), 25 °C (circles), and 35 °C (squares). The lines are the best fits from nonlinear regression analysis according to the linear extrapolation model (LEM). (B) Temperature dependence of the *m*-values obtained from LEM analysis. The regression line has a slope dm/dT = 0.066 ± 0.007 kJ mol^{-1} M^{-1} K^{-1}.

Indeed, as shown illustrated in Figure 1 ΔH_{WA} simulates perfectly the experimental trace.

(B) Urea-Induced Unfolding. The stability of PDZ2 at 5, 15, 25, and 35 °C was assessed from isothermal urea-induced unfolding experiments by following the change in ellipticity at 219 nm. The data could be modeled with a two-state transition between native and unfolded protein (Figure 2). Both the midpoints of transition, $U_{1/2}$, and the equilibrium *m*-value, $m_{\mathrm{eq}} = -\mathrm{d}\Delta G_{\mathrm{urea}}/\mathrm{d}[\mathrm{urea}]$, depend on the temperature with slopes $\mathrm{d}U_{1/2}/\mathrm{d}T = -0.040 \pm 0.008$ M K^{-1} and $\mathrm{d}m_{\mathrm{eq}}/\mathrm{d}T = 0.066 \pm 0.007$ kJ mol^{-1} M^{-1} K^{-1}, respectively. The mean m_{eq} (∼6 kJ mol^{-1} M^{-1}) corresponds very well with the prediction of statistical analysis correlating changes of solvent-accessible surface with the magnitude of the experimental *m*-value (*52*). The unfolding free energy estimates at zero urea concentration, $\Delta G_{\mathrm{U}}(\mathrm{W})$, calculated according to the linear extrapolation procedure are listed in Table 1.

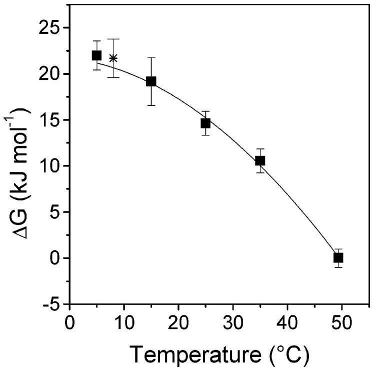


FIGURE 3: Stability curve of PDZ2. The unfolding free energies obtained by isothermal urea-induced unfolding between 5 and 35 °C were combined with $\Delta G_{\mathrm{U}} = 0$ at 49.4 °C (the midpoint of thermal unfolding), and the Gibbs−Helmholtz equation was fit to the data to optimize the values for T_{m}, ΔH_{m}, and ΔC_p. The continuous line represents the best fit obtained with the following parameters: $T_{\mathrm{m}} = 49.5 \pm 1.0$ °C, $\Delta H_{\mathrm{m}} = 255 \pm 30$ kJ mol^{-1}, and $\Delta C_p = 4.36 \pm 1.90$ kJ K^{-1} mol^{-1}. The asterisks represent ΔG_{U} obtained from GdmCl-induced unfolding followed by fluorescence at 8 °C (*17*)

(C) Stability Curve of PDZ2. The Gibbs−Helmholz equation (eq 2) was fit to the combined data from thermal and urea-induced unfolding to calculate the PDZ2 stability curve (Figure 3). The best fit was obtained with the following parameters: $T_{\mathrm{m,fit}} = 49.5 \pm 1.0$ °C, $\Delta H_{\mathrm{m,fit}} = 255 \pm 30$ kJ mol^{-1}, and $\Delta C_{p,\mathrm{fit}} = 4.35 \pm 1.90$ kJ K^{-1} mol^{-1} (Table 1). The correspondence between the ΔC_p and ΔH values obtained by considering data in a broad temperature range and ΔC_p and ΔH_{WA} measured directly by DSC is excellent. The only one previous estimate of PDZ2 stability is 21.7 ± 2.1 kJ mol^{-1} at 8 °C and was obtained by GdmCl-induced denaturation followed by the intrinsic changes in PDZ2 fluorescence (*17*). As shown in Figure 3, the agreement of this value with the calculated stability curve is very good. The coincidence of ΔG_{U} measured by three different methods is a very strong argument that PDZ2 unfolding closely obeys the two-state model. The fraction of unfolded protein is less than 1% between 5 and 30 °C. ITC experiments were performed in this temperature range.

Energetics of RG and APC Binding to PDZ2. (A) Thermodynamic Parameters Deduced from ITC. Addition of aliquots of the RG peptide to PDZ2 in the calorimetric cell produces measurable heat effects, which saturate as the

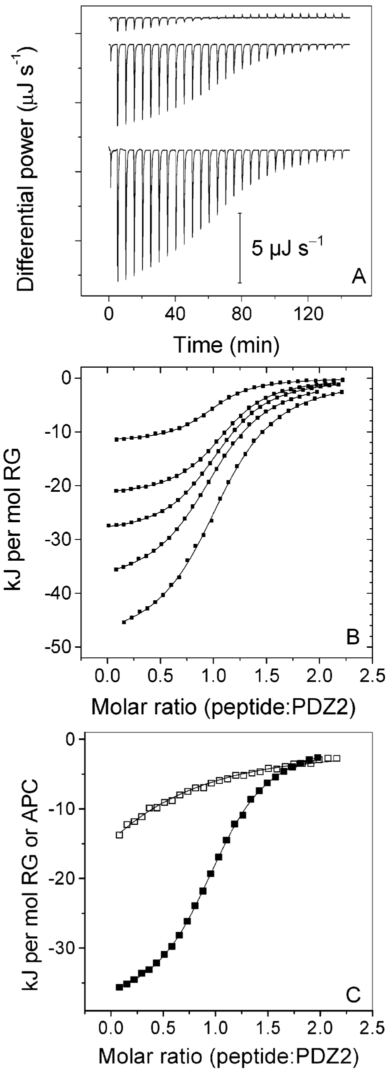


FIGURE 4: ITC binding experiments. All experiments were performed in 50 mM sodium phosphate and 150 mM NaCl, pH 6.8. The PDZ concentration in the cell was 30−80 μM; the peptide concentration in the syringe was 400−1000 μM. (A) Raw calorimetric output from experiments with the RG peptide. Experiments at 5, 18, and 30 °C are shown (top to bottom). The thermograms were shifted on the *Y*-axis for clarity. (B) Binding isotherms describing formation of the PDZ2/RG complex at 5, 12, 18, 25, and 30 °C (top to bottom). Symbols represent the integrated heats after normalization for the molar concentration. Continuous lines are nonlinear fits for a 1:1 binding model. (C) Comparison of the binding isotherms obtained at 25 °C with peptide APC (top trace and open symbols) and RG (bottom trace and filled symbols).

molar ratio of peptide to PDZ2 increases (Figure 4). From the shape of the titration curve, the binding stoichiometry, *n*, the association constant, K_A, and the apparent calorimetric enthalpy of association, ΔH_A, were calculated. The stoichiometry was 0.98 ± 0.10 (mean ± SD of 15 experiments), in agreement with a 1:1 complex seen in the NMR structure. The association constant at 25 °C is $(1.94 \pm 0.17) \times 10^5$ M^{-1}. The corresponding dissociation constant is thus ~5 μM. This first, direct, in-solution equilibrium measurement of the binding affinity of PDZ2 for RG C-terminal sequences confirms the results of two previous studies, which estimated the dissociation constant of the PDZ2/RG complex as being lower than 30 μM based on chemical shifts and amide exchange rates followed by NMR spectroscopy (*10*, *18*). From the rates of association and dissociation of the C-terminal RA-GEF2 pentapeptide to and from the PDZ2 Y43W mutant, a $K_d = 6.5 \pm 1.5$ μM was calculated (*53*). To our knowledge, there are three published studies in which binding of C-terminal sequences to PDZ was characterized by titration calorimetry (*19*, *21*, *54*). The third PDZ domain (class I) of the PSD-95 protein binds the hexapeptide KKETEV with $K_d \sim 2$ μM (*54*). Dissociation constants in the range 200 nM to 50 μM characterize binding of short peptides to full-length syntenin or its PDZ tandem (*19*). An ELKS1b-derived peptide binds with high affinity ($K_d = 270$ nM) to the RIM1 PDZ domain (*21*).

PDZ2 was reported to bind the C-terminal stretch of the APC protein with low nanomolar affinity (*14*). We performed ITC experiments with a synthetic peptide encompassing the last 19 amino acids of the APC protein. Figure 4C compares experiments with RG and APC peptides at identical conditions. Much to our surprise, the obtained isotherms with APC are compatible with a K_d above 50 μM. The reason for this striking discrepancy with the results of Erdmann et al. is not clear at present. The vast majority of data indicating low micromolar to nanomolar affinities of PDZ domains to other proteins was collected by surface plasmon resonance experiments, and K_d's were calculated from the rates of association and dissociation (*15*, *16*, *55*). It has been demonstrated that rebinding to the sensor surface might lead to understimates of the genuine dissociation rate constant and might thus cause an overestimation of the binding affinity (*56*, *57*). In particular, next to high surface coverage, fast association rates might make rebinding very effective. Fast association is expected for ligands with relatively low steric requirements, and this is possibly the case for peptide binding to PDZ domains, since the intermolecular interface is relatively small. On the other hand, the experiments reported by Erdmann et al. were performed with a GST-PDZ2 at pH 7.4 in the presence of small amounts of nonionic detergent, and direct comparison between the two studies has to be done with caution. Nevertheless, our data demonstrate that the RA-GEF2-derived peptide binds much more strongly to isolated PDZ2 than the APC-derived peptide under identical experimental conditions. It remains to be seen in how far the differences in affinity can be related to the sequential differences at position −1 (Ala in RG versus Ser in APC) and beyond position −3, where peptide library studies indicated a preference for polar and (predominantly) negatively charged side chains (*16*). This pattern is clearly evident in RG while positions −4 and −5 are occupied by leucine and tyrosine in APC. In the context of the cell, the strength of the PDZ2/APC interaction may turn to be an important issue, since APC binding to the hPTP1E protein appears to modulate the steady-state levels of tyrosine phosphorylation of proteins playing a role in the regulation of cell division, migration, and cell adhesion (*14*).

The thermodynamic parameters characterizing the PDZ2/RG interactions are summarized in Table 2 and are presented in graphical form in Figure 5. K_A exhibits a small but statistically significant temperature variation, decreasing from $(4.70 \pm 0.84) \times 10^5$ M^{-1} at 5 °C to $(1.80 \pm 0.23) \times 10^5$ M^{-1} at 30 °C. The free energy of association is almost

Table 2: Thermodynamic Parameters Characterizing RG Binding to PDZ2 Measured by ITC[a]

T (°C)	$K_A \times 10^{-5}$ (M^{-1})	ΔG (kJ mol^{-1})	ΔH (kJ mol^{-1})	$T\Delta S$ (kJ mol^{-1})	ΔS (J K^{-1} mol^{-1})
5	4.70 ± 0.84	−30.2 ± 0.2	−11.7 ± 0.7	18.5 ± 0.2	66.5 ± 0.7
12	3.04 ± 0.23	−29.9 ± 0.1	−20.6 ± 1.4	9.3 ± 1.3	32.7 ± 0.4
18	2.63 ± 0.26	−30.2 ± 0.1	−29.1 ± 1.6	1.1 ± 1.6	3.8 ± 5.0
25	1.94 ± 1.69	−30.2 ± 0.1	−39.6 ± 0.9	−9.4 ± 0.9	−31.6 ± 3.0
25[b]	1.63 ± 0.30	−29.8 ± 0.2	−40.2 ± 3.0	−10.5 ± 3.0	−35.1 ± 17.0
25[c]	1.60 ± 0.20	−29.7 ± 0.2	−38.9 ± 2.0	−9.2 ± 2.0	−30.9 ± 6.7
30	1.80 ± 0.23	−30.5 ± 0.1	−47.3 ± 2.2	−16.8 ± 2.2	−55.4 ± 7.0
55[d]	0.18 ± 0.4	−26.8 ± 50	−80 ± 40	−53.2 ± 30.0	−162 ± 100

[a] All experiments were performed in 50 mM sodium phosphate and 150 mM NaCl, pH 6.8, unless otherwise indicated. [b] Experiment in MES buffer. [c] Experiment in Bis-Tris buffer. [d] Deduced from DSC experiments as described in the text.

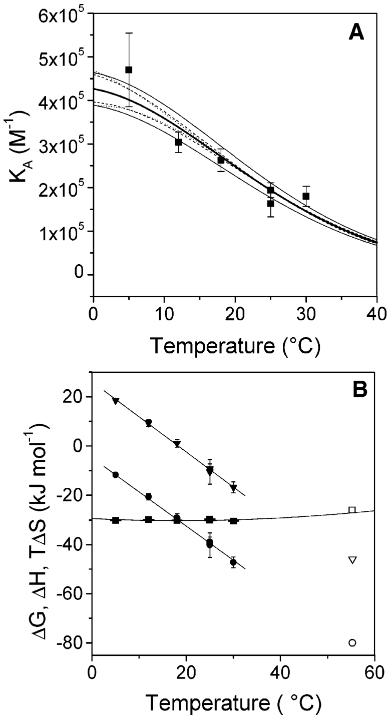


FIGURE 5: Energetics of RG binding to PDZ2 measured by ITC. All numerical values are listed in Table 2. (A) Temperature dependence of the association constant, K_A. Symbols represent the experimental data. The thick continuous line is the calculated temperature dependence of K_A, calculated according to eq 9 using as reference the data at 25 °C: $K_A = 1.94 \times 10^5$ M^{-1}, $\Delta H_A = -39.6$ kJ mol^{-1}, and d$C_{p,A} = -1.4$ kJ K^{-1} mol^{-1}. The pairs of lines illustrate the influence of the experimental uncertainty (as upper and lower limits) on the calculated temperature variation of K_A. Key: thin continuous lines, $K_A \pm 1.69 \times 10^4$ M^{-1} at fixed ΔH_A and $\Delta C_{p,A}$; dotted lines, $\Delta C_{p,A} \pm 0.2$ kJ K^{-1} mol^{-1} at fixed K_A and ΔH_A; dashed lines, $\Delta H_A \pm 2$ kJ mol^{-1} at fixed K_A and $\Delta C_{p,A}$. (B) Temperature variation of ΔG_A (squares), ΔH_A (circles), and $T\Delta S_A$ (triangles). The continuous lines associated with ΔH_A and $T\Delta S_A$ data points are linear regression lines. The line associated with the ΔG_A data points was calculated as $\Delta G_A = -RT \ln K_A$ using the simulated temperature variation of K_A (thick line in panel A). The open symbols are ΔH_A and $T\Delta S_A$ derived from DSC experiments at 55 °C.

constant in the studied range of temperatures. The apparent enthalpy changes, ΔH_A, and entropy changes, $T\Delta S_A$, of association vary with temperature. Complex formation is exothermic above 0 °C. The same ΔH_A was measured in MES and Bis-Tris buffers having different heats of protonation. It follows that there is no (net) proton release/uptake accompanying binding. The same conclusion was reached also by Saro et al. for the PSD-95 protein PDZ3 domain, another member of class I (*54*). A pK_a shift of ~1 pH unit upon binding was reported earlier for His 71, which is located at the PDZ2 binding site and forms a hydrogen bond with the hydroxyl group of Ser −2 of a Fas-derived peptide (*17*). This interaction is canonical for class I PDZ domains and, not surprisingly, is seen also in the PDZ2/RG structure. However, the measured lack of ΔH_A variation with the buffering species implies that His 71 is deprotonated both in free PDZ2 and in the PDZ2/RG complex. To further clarify the problem, we calculated the pK_a's of titratable groups in the free and peptide-bound PDZ2. The pK_a of His −71 is very low, 2.4 ± 1.2, reflecting the almost complete burial of the side chain upon peptide binding and the strong interaction with the hydroxyl group of Ser −2 of the incoming peptide. However, the pK_a of His −71 is downshifted also in the free protein. Depending on the choice of dielectric constant describing the dielectric properties of the protein, the calculated pK_a's of the unbound form vary between ~4 ($\epsilon = 4$) and ~5 ($\epsilon = 20$). The downward shift of the pK_a of the unbound form relative to the model pK_a (6.6 in this implementation) appears to be dominated by the desolvation term and interactions with permanent charges of the protein multipole and, to a lesser extent, by interactions with other titratable sites. As a control, we checked the pK_a's of the three other histidine side chains, which are not involved in interactions with RG. They are all in the range 6−7 and, most importantly, are not influenced by the presence of the peptide in the binding pocket. The calculations suggest that binding at pH 6.8 is not accompanied by proton release, in accord with the available experimental evidence.

The entropy of complex formation is positive below about 18 °C and negative at higher temperature (Figure 5). Hence, binding is favored enthalpically at all temperatures, whereas it is opposed by entropy above 18 °C. The temperature variation of the binding enthalpy, dΔH_A/dT, represents the heat capacity change of association, $\Delta C_{p,A}$. Linear regression of the data shown in Figure 5 yields $\Delta C_{p,A}$ of −1.44 ± 0.20 kJ K^{-1} mol^{-1}.

(B) DSC Studies. Further details on the binding energetics can be extracted from DSC experiments. The temperature-induced conformational transition of the preformed PDZ2/RG complex is highly reversible. The partial molar heat capacity function is shown in Figure 6 together with the traces recorded for the isolated components. The heat capacity of the RG peptide changes negligibly upon heating and is very close to the heat capacity calculated for an unstructured, completely hydrated peptide with the RG sequence. Melting of the complex produces a single heat

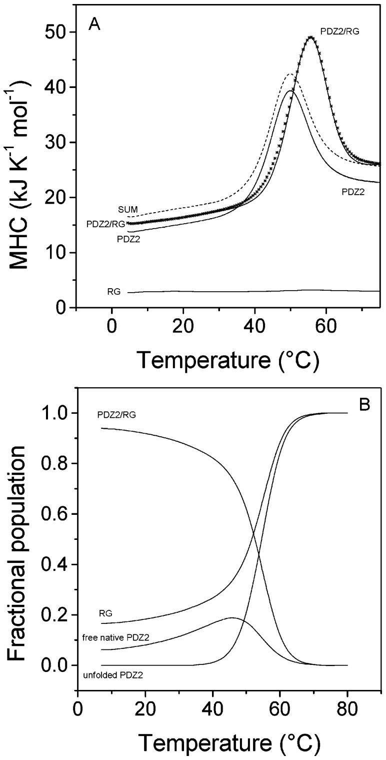


FIGURE 6: Heat capacity curves of the PDZ2/RG complex and the isolated components. (A) Thermograms recorded in standard buffer, 50 mM sodium phosphate and 150 mM NaCl, pH 6.8, at a heating rate of 1 deg min^{-1}. The labels associated with the heat capacity functions indicate experiments with the isolated RG peptide (trace RG; concentration was 250 μM), isolated protein (trace PDZ2; concentration was 222 μM), and a mixture of protein and peptide (trace PDZ2/RG shown with symbols; concentrations were 222 μM protein and 250 μM peptide). For experimental reasons, the highest concentration in DSC experiments was ~0.5 mg mL^{-1}, and precise determination of the absolute heat capacity was not possible. However, the deviation from the predicted heat capacity was only ~0.5 kJ mol^{-1} K^{-1}. The trace was therefore shifted as to match the calculated heat capacity at 75 °C. The trace labeled SUM (dashed line) is the algebraic sum of the heat capacities of the isolated components; i.e., SUM = RG + PDZ2. This is the expected heat capacity if the system was in the hypothetical dissociated state at all temperatures. The continuous line associated with experiment PDZ2/RG visualizes the result of simulation according to eqs 3−14. The curve is defined by the following parameters: K_A(25 °C) = 2.2×10^5 M^{-1}, ΔH_A(25 °C) = −35.5 kJ mol^{-1}, $\Delta C_{p,A}$ = −1.7 kJ K^{-1} mol^{-1}, $\Delta H_{m,P}$(49.4 °C) = 260 kJ mol^{-1}, and $\Delta C_{p,P}$ = 4.8 kJ K^{-1} mol^{-1}. (B) Calculated temperature-induced changes of the fractional population of unbound and native PDZ2, unbound and unfolded PDZ2, unbound peptide RG, and the PDZ2/GR complex.

absorption peak whose temperature of maximum heat absorption (which can be regarded as the apparent transition temperature or melting temperature, $T_{m,C}$) is higher by ~6 °C than the melting temperature of PDZ2 alone. Hence, PDZ2 is stabilized in the complex. The dissociation of the complex and the concurrent unfolding of the protein are taking place within a relatively narrow temperature interval.

What can be learned from the thermogram? Figure 6 shows the expected heat capacity of the system in a hypothetical state without intermolecular interactions. This function, C_p^{SUM}, was calculated by simple algebraic summation of the heat capacities of free PDZ2 and free RG; i.e., $C_p^{SUM} = C_p^{PDZ2} + C_p^{RG}$. Between 5 and 30 °C C_p^{SUM} is higher than the heat capacity of the (partly) associated state, $C_p^{PDZ2/RG}$. This reflects the negative heat capacity change accompanying binding. The difference $C_p^{PDZ2/RG} - C_p^{SUM}$ is ~−1.6 ± 0.15 kJ K^{-1} mol^{-1} on average and is thus very close to $\Delta C_{p,A}$ obtained by ITC. However, the close correspondence of the two numbers is perhaps a fortuitous coincidence in view of the uncertainties in determination of the absolute heat capacities at relatively low concentrations and the use of calculated partial specific volumes. Much more important for the following discussion is the fact that the slopes of $C_p^{PDZ2/RG}$ and C_p^{SUM} are very close to each other: 96 ± 5 and 112 ± 7 J K^{-2} mol^{-1}, respectively (Table 3). These values imply, in principle, a temperature-dependent $\Delta C_{p,A}$. However, the integral $\int_T (C_p^{PDZ2/RG} - C_p^{SUM})\,dT$ is a small number, and the resulting deviation of the dH_A/dT plot from linearity is certainly lower than the accuracy of the measurements. It follows that the formation of the PDZ2/RG complex is not linked to detectable temperature-dependent structural changes, or at least such changes contribute to the enthalpy of association but do not induce significant temperature dependence of ΔC_p.

The apparent calorimetric enthalpy of complex melting and dissociation, $\Delta H_{cal,PDZ2/RG}$, can be obtained by integration of the heat absorption peak above the chemical baseline constructed in the usual way and amounts to 370 ± 30 kJ mol^{-1} (mean ± SD of three experiments). Integration in respect to ln$T$ yields the apparent entropy of the process, $\Delta S_{cal,PDZ2/RG}$ = 1.07 ± 0.09 kJ K^{-1} mol^{-1}. The apparent calorimetric heat capacity change of complex melting and dissociation, $\Delta C_{p,PDZ/RG}$ corresponds to the difference between the pretransitional and posttransitional portions of the trace linearly extrapolated at the apparent melting temperature; it amounts to 6.0 ± 0.8 kJ K^{-1} mol^{-1}. The difficulty in giving a straightforward interpretation of these parameter numbers as representing the sum of protein unfolding and complex dissociation arises from the fact that the distribution of molecular species in the transition zone is not known a priori. To this end, the excess heat capacity function was analyzed by a model which takes into account the temperature-induced changes in the concentration of free and bound native protein, unfolded protein, and free and bound peptide. This analysis combines the energetics of folding of PDZ2 with the energetics of complex formation described above. The deconvolution technique was developed more than 10 years ago (*33*) and was applied to different types of protein−ligand interactions (*32*, *58*−*62*). Figure 6 presents the results of simulations according to the combined eqs 3−14 as outlined in Experimental Procedures. The experimental excess heat capacity profile of the PDZ2/RG mixture is reproduced with parameters that are in good agreement with parameters describing melting of PDZ2 and PDZ2/RG complex formation at 25 °C. Figure 6B shows the calculated population of each molecular species. At the selected concentrations (~250 μM), the system is partly dissociated at low temperatures, and the concentration of the complex gradually decreases upon heating. PDZ2 melting and complex dissociation are tightly coupled. In the transition region, melting of the complex releases small amounts of free folded protein. This is in accord with a reaction order between 1

Table 3: Thermodynamic Parameters Describing Thermal Melting of PDZ2, RG, and the PDZ2/RG Complex Measured by DSC[a]

	T_m (°C)[b]	ΔH_{cal} (kJ mol^{-1})	ΔS_{cal} (J K^{-1} mol^{-1})	ΔC_p (kJ K^{-1} mol^{-1})	dC_p/dT (J K^{-2} mol^{-1})
PDZ2	49.4	273 ± 10	846 ± 30	4.45 ± 0.20	107 ± 5
	55.2	300 ± 12[c]	925 ± 35[c]		
RG		0	0	0	5 ± 5
PDZ2/RG	55.2	370 ± 30	1070 ± 90	6.00 ± 0.60	96 ± 5
PDZ2 + RG					112 ± 7
complex − components	55.2	70 ± 32	145 ± 95	1.55 ± 0.65	−16 ± 8
complex dissociation	55.2	80 ± 40	165 ± 100	1.76 ± 0.70	−18 ± 9

[a] All experiments were performed in 50 mM sodium phosphate and 150 mM NaCl, pH 6.8. [b] Corresponds to the maximum of the C_p/T curve. [c] Parameters calculated by extrapolation.

and 2. Indeed, the reaction order best describing the melting trace is 1.85 (not shown).

With help of the known molar fractions as a function of the temperature (Figure 6B), we can now deconvolute the total enthalpy change ($\Delta H_{cal,PDZ2/RG}$), entropy change ($\Delta S_{cal,PDZ2/RG}$), and heat capacity change ($\Delta C_{p,PDZ2/RG}$) measured upon melting of the PDZ2/RG mixture. The contribution of the peptide to any parameter can be neglected since temperature-induced conformational changes are calorimetrically silent (see the essentially temperature-independent heat capacity function in Figure 6A). The following equations hold at $T_{m,C}$:

$$\Delta H_{cal,PDZ2/RG} = \Delta H_{cal,P}^{T_{mC}} - f_C \Delta H_{cal,A}^{T_{mC}}$$

$$\Delta S_{cal,PDZ2/RG} = \Delta S_{cal,P}^{T_{mC}} - f_C \Delta S_{cal,A}^{T_{mC}}$$

$$\Delta C_{p,PDZ2/RG} = \Delta C_{p,P} - f_C \Delta C_{p,A}^{T_{mC}}$$

The first terms on the right-hand side are the parameters describing PDZ2 unfolding at $T_{m,C}$ and are calculated as $\Delta H_{cal,P}^{T_{mC}} = \Delta H_{cal,P} + \Delta C_{p,P}(T_{m,C} - T_{m,P})$ and $\Delta S_{cal,P}^{T_{mC}} = \Delta S_{cal,P} + \Delta C_{p,P} \ln(T_{m,C}/T_{m,P})$; $\Delta C_{p,P}$ is a constant over the small temperature range of extrapolation. The second terms on the right-hand side are the energetic parameters of the PDZ/RG complex. The minus sign accounts for complex dissociation. The multiplier f_C represents the molar fraction of the complex at 30 °C, i.e., at the onset of the heat absorption peak, and equals 0.89. For the only unknowns in the above equations, one obtains $\Delta H_{cal,A}^{T_{mC}} = -80 \pm 40$ kJ mol^{-1}, $\Delta S_{cal,A}^{T_{mC}} = -165 \pm 100$ kJ K^{-1} mol^{-1}, and $\Delta C_{p,A} = -1.8 \pm 0.8$ kJ K^{-1} mol^{-1}. The free energy change is thus -26 kJ mol^{-1}. These are the thermodynamic parameters of association of the PDZ2/RG complex at 55 °C (open symbols in Figure 5). Although the rules of error propagation place a large uncertainty of the calculated numbers, the correspondence between the mean values of the thermodynamic functions obtained by ITC at lower temperatures and by DSC at 55 °C is surprisingly good in view of the crude approximations underlying the calculations. First, any temperature dependence of ΔC_p is neglected. Second, both molecules are considered to be structurally and energetically invariant in the entire temperature range. Third, there is no robust way to justify the validity of the two-state model, and any possible redistribution of binding modes in the transition region is ignored. Fourth, the implicit assumption is that there are no interactions taking place between the peptide and the protein in its denatured state. Nevertheless, the success of the modeling (based on van't Hoff formalism) and the deduced calorimetric enthalpy, entropy, and heat capacity changes indicate that the thermodynamic parameters obtained by ITC and DSC describe the thermodynamics of RG binding to PDZ2 in a consistent way and that the procedure is a valuable alternative to obtain thermodynamic information about protein−ligand binding at temperatures beyond the usually limited temperature range covered by ITC experiments. More importantly, the procedure allows a reliable interpolation of the binding constant to the physiologically relevant temperature of 37 °C.

Energetic Partitioning of the Binding Affinity. The driving force of RG binding to PDZ2 at 25 °C (and in physiologically relevant temperature range) is the favorable enthalpy change. The same is true for PDZ3 of PSD-95, which is also a class I PDZ domain (*54*). In contrast, favorable entropy change dominates the affinity of peptide binding to syntenin PDZ1 and PDZ2 domains (*19*). The latter two domains exhibit degenerate specificity and bind peptide ligands of classes I, II, and III. Neither of them has a histidine (typical for class I) or a hydrophobic residue (typical for class II) or a tyrosine (typical for class III) at the beginning of helix α2. With the limited data available, it is impossible to interpret or generalize the different enthalpy−entropy signature of class I-specific complexes, as compared to PDZ−peptide complexes with degenerate specificity. From a structural perspective, the large spread of binding enthalpies (−2 to −40 kJ mol^{-1} at 25 °C) and the apparent effective enthalpy/entropy compensation are surprising. First, the interactions anchoring the very C-terminal hydrophobic residue are ubiquitous and conservative. Second, it appears that backbone−backbone hydrogen bonding contributes essentially to the binding affinity in a "nonspecific" manner. Although progress has been made toward prediction of binding affinities and actual design of ligands capable of binding PDZ domains (*63*, *64*), the emergent compatibility of the PDZ binding site with ligands that bind, driven by enthalpy or entropy, or both, could turn out to be an obstacle in the rational optimization of lead compounds for high affinity *and* specificity (*65*).

Structural Parametrization of Energetic Changes. However good the precision of measured thermodynamic parameters might be, they reflect the total energetic changes of the considered system. The challenging problem is indeed to relate energy terms to structural features in order to understand the molecular basis of the binding affinity. Starting from the early 1990s, different approaches have been followed (*41*, *66*−*70*). Among different parametrization schemes, we use here the one developed by Murphy and Freire, which has been extensively tested against experimental data on unfolding and binding and enjoys commensurate popularity (*41*, *42*, *69*). For brevity we henceforth refer to the procedure as M&F (standing for Murphy and Freire; the contribution of numerous members of the Murphy and Freire laboratories to the refinement of the calculation is implicitly acknowledged).

Since the M&F calculations rely on solvent accessibility differences, the structures of the binary complex and its components in their free state should be known at high resolution or should be carefully modeled. The NMR structure of the PDZ2/RG complex was solved [1D5G (*10*)]. The NMR structure of PDZ2 in the unbound form is also available [3PDZ (*11*)]. The five C-terminal residues of RG are ordered in the binding site, yet free RG obtains a random coil conformation according to CD measurements, and therefore, the RG structure extracted from the complex is inadequate to model the free RG conformation. To overcome the problem and, more importantly, to create a consistent set of structures, we performed MD simulations of free PDZ2, free RG, and the PDZ2/RG complex in explicit water. According to the usual criteria (C_α RMSD, radius of gyration, solute−solute and solute−solvent energy terms, intermolecular distances) the MD trajectories were well equilibrated to serve as a model of the dynamic behavior of the complex and its components (not shown). The average NMR and MD structures of the PDZ2/RG complex and its components are very close to each other, the C_α RMSD (secondary structure elements) being between 0.87 and 1.36 Å in cross-comparisons. However, we noted a slight structure contraction of the complex during MD. The change in ASA is relatively small but significant in comparison with the spread of ASA within the NMR ensemble. Therefore, we tested the calculations against both ensembles according to the following reaction schemes:

$$\begin{array}{l}\text{free PDZ}\\ \text{PDZ}_{\text{3PDZ}}\text{: model NMR}_{\text{conf}}\\ \text{PDZ}_{\text{free,MD}}\text{: model MD}_{\text{conf}}\end{array} + \begin{array}{l}\text{free RG}\\ \text{RG}_{\text{free,MD}}\text{: models NMR}\\ \text{RG}_{\text{free,MD}}\text{: models MD}\end{array} \xrightarrow[\text{models conf}]{1} \begin{array}{l}\text{PDZ/RG}\\ \text{PDZ/RG}_{\text{1D5G}}\text{: models NMR}\\ \text{PDZ/RG}_{\text{MD}}\text{: models MD}\end{array}$$

$$\begin{array}{l}\text{“binding-competent” PDZ}\\ \text{PDZ}_{\text{1D5G}}\text{: model NMR}_{\text{rb}}\\ \text{PDZ}_{\text{bound,MD}}\text{: model MD}_{\text{rb}}\end{array} + \begin{array}{l}\text{free RG}\\ \text{RG}_{\text{free,MD}}\text{: models NMR}\\ \text{RG}_{\text{free,MD}}\text{: models MD}\end{array} \xrightarrow[\text{models rb}]{2} \begin{array}{l}\text{PDZ/RG}\\ \text{PDZ/RG}_{\text{1D5G}}\text{: models NMR}\\ \text{PDZ/RG}_{\text{MD}}\text{: models MD}\end{array}$$

Reaction 1 represents the real process, taking into account the conformational changes of the protein domain accompanying binding, while reaction 2 is a hypothetical process, where binding is approximated as a rigid-body association. In model NMR, ensemble 3PDZ (human PDZ2) was used to represent the free protein, while the binding-competent conformation was extracted from the structure of the human PDZ2/RG complex (1D5G), which serves also as a model of the associated state of the system. In model MD, the structures of the free and binding-competent PDZ2 were taken from the trajectories of the free protein and of the complex, respectively. In both models, the free peptide was modeled with the MD structure. The calculated surfaces represent averages over the corresponding ensembles.

The calculated parameters are listed in Table 4. The comparison with the experimental parameters is done at 25 and 60 °C for ΔH_A and at 18 °C for ΔS_A, where the total entropy contribution is zero. Model NMR_{conf} predicts the binding enthalpy within the range of the other model's estimates, yet all other terms are significantly different and completely contradict the experimental trend. As noted by Walma et al., the free PDZ2 structure suffers significant problems according to several established structure quality tests (*71*). However, no improvement was achieved by using the NMR structure of the mouse, highly homologous PDZ2 domain (1GM1). The failure of the NMR_{conf} model reiterates that the quality of structures is critical in ASA-based parametrization and warns against the use of homologous structures. Since model NMR_{rb} predicts endothermic binding at 25 °C, we focus the following discussion on the results obtained with the MD ensembles.

Table 4: Calculated and Experimental Energetics of RG Binding to PDZ[a]

	NMR_{rb}[b]	NMR_{conf}[c]	MD_{rb}[d]	MD_{conf}[e]	exp
ΔC_p (kJ K^{-1} mol^{-1})	−0.56	0.16	−0.65	−0.76	−1.44
ΔH at 25 °C (kJ mol^{-1})	8	−7	−9	−15	−40
ΔH at 60 °C (kJ mol^{-1})	−11	−2	−31	−41	−88
ΔS at 18 °C (J K^{-1} mol^{-1})	15	nd	45	80	0
total dehydration part	185	−45	245	300	
total conformational part	−135	nd	−165	−185	
rotation/translation	−35	nd	−35	−35	

[a] In all calculations the MD ensemble of RG was used as the model of the unbound state of the peptide. [b] With the NMR ensemble extracted from the human PDZ/GR complex (1D5G) as the model of the unbound protein. [c] With the NMR ensemble of the free human PDZ2 (3PDZ) as the model of the unbound protein. [d] With the MD ensemble of PDZ2 extracted from the simulation of the PDZ2/GR complex. [e] With the MD ensemble of PDZ2 simulated in the unbound state.

Model MD_{conf} is closest to the experiment in terms of ΔH and ΔC_p, indicating that the transition from the free to the binding-competent state is likely to make an energetic contribution to binding. Comparison of the numbers listed in Table 4 leads to the conclusion that the M&F model underestimates $\Delta C_{p,\text{A}}$ and ΔH_{A} at all temperatures and predicts a positive ΔS_{A} at 18 °C. The magnitude of the absolute deviations from the experiment, ~0.5−0.7 kJ K^{-1} mol^{-1} ($\Delta C_{p,\text{A}}$), 30−35 kJ mol^{-1} (ΔH_{A} at 25 °C), and 40−80 J K^{-1} mol^{-1} (ΔS_{A} at 18 °C), could be considered as representing the inherent inaccuracy of structure-based predictions of the enthalpy and heat capacity changes in protein unfolding (both large in comparison to binding parameters), when applied to a large set of data (experimental values taken from ref *50*). On the other hand, however, the enthalpy of binding of small ligands and protein−protein interactions could be predicted with very high precision (*69*, *72*, *73*). Noteworthy, the parametrization scheme developed by Makhatadze and Privalov (*49*) using different ideas and methodology yields essentially the same overall picture: it predicts $\Delta C_{p,\text{A}}$ within 0.6−0.9 kJ K^{-1} mol^{-1} and ΔH_{A} within −10 to −20 kJ mol^{-1} at 25 °C and significantly overestimates the temperature where ΔS_{A} crosses zero. Hence, it appears that the experimental parameters "in excess" over the structural prediction are caused by some physical process intimately coupled to binding. Where could the observed discrepancies stem from? A couple of possibilities are discussed below.

(A) Heat Capacity Change. Experimental ΔC_p values more negative than those predicted from surface burial have been reported for a variety of systems. Apart from the "trivial" case, where large refolding transitions accompanying binding contribute to the apparent ΔC_p change, different explanations have been forwarded with a common theme: The changes in hydration are not the only physical source of ΔC_p decrease concomitant with binding. Additional contributions have been ascribed to (i) small structural perturbations leading to redistribution of easily excitable vibrational modes (*74*), (ii) ligand-induced narrowing of the distribution of enthalpic microstates (*75*), (iii) formation of highly cooperative arrays of noncovalent bonds (*76*), (iv) unusually large temperature dependence of the intrinsic heat capacities of the associated and dissociated states of the system (*77*), and (v) entrapment of water at the binding interface (*78*, *79*). Propositions i−iii represent general concepts that are difficult to verify experimentally. We find no experimental support for proposition iv because the heat capacity of the complex and the sum of the heat capacities of RG and PDZ2 change almost in parallel with temperature (Table 3 and Figure 6). Significant contribution from water entrapment is also unlikely since the contribution of a trapped water molecule to ΔC_p was estimated as being lower than -40 J K^{-1} mol^{-1} in an nonpolar environment (*79*, *80*) and could probably not exceed -75 J K^{-1} mol^{-1} (the heat capacity of bulk water). One to three waters have been detected so far in PDZ−peptide complexes by crystallography, and the residence time of each individual water molecule observed in our MD simulations was never longer than 3.5 ns.[2]

(B) Enthalpy Change. We observe no measurable net heats from protonation/deprotonation events. Also, no ions are present at the protein−peptide interface on the time scale of the MD simulations. The enthalpic effect of the water molecules found at or close to the interface (one on average over the 15 ns MD trajectory) cannot be sizable (*81*, *82*). Hence, the structure-based prediction underestimates the generic binding enthalpy (*41*, *42*). It should be noted that eq 16 relates the enthalpy (of unfolding) to the average packing density of proteins and assumes no residual enthalpy after breakage of all relevant noncovalent bonds. We have estimated the packing density by calculating the energy-weighted distance average between atom−atom pairs which form the typical array of noncovalent bonds in proteins and protein−protein complexes (*81*). The obtained values are much larger than those for the reference set analyzed in ref *81* and, hence, indicate substantially looser packing of the PDZ2/RG interface in comparison with the interior of an average protein. Furthermore, the hydrogen bonds at the interface are not particularly short or geometrically optimized, or networked, and visual inspection does not identify other polar interactions that might explain a strong enthalpic effect. We conclude that the experimental ΔH_A contains a contribution of an exothermic process that is not related to the formation of intermolecular contacts.

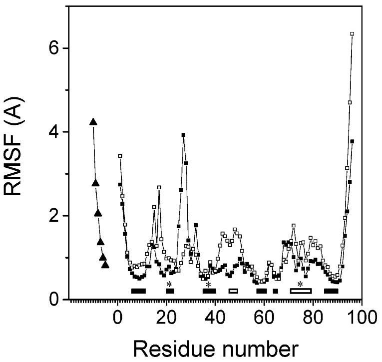


FIGURE 7: Root mean square fluctuations (RMSF) of the C_α atoms around the average MD structures. Shown is RMSF of unbound PDZ2 (open symbols), bound PDZ2 (filled symbols), and bound RG (triangles). Only RMSF of the six C-terminal residues of RG (positions -5 to 0; left to right) are plotted. Secondary structure elements are represented with filled bars (β-strands) and open bars (α-helices). The asterisks indicate secondary structure elements that form the peptide binding site (strands $\beta 2$ and $\beta 3$ and α-helix 2; left to right).

(C) Entropy Changes. At 18° ΔS_A is zero and, therefore, the total dehydration contribution is precisely balanced by the total conformational entropy change plus the loss of rotational/translational degrees of freedom (eq 17). The latter term is still a matter of much debate. Following M&F for consistency, we use here -35 J K^{-1} mol^{-1}, which is close to the cratic entropy and has been found to describe reasonably the loss of rotational/translational degrees of freedom in diverse experimental systems (*44*−*46*). The M&F model predicts "excess" positive entropy (favoring binding). In fact, the calculated conformational entropy (disfavoring binding) is perhaps an overestimate. In the calculation of the conformational contribution we have conservatively assigned the ΔS_{bb} and ΔS_{ex-u} terms to all five peptide residues that appear structured in the binding pocket. The entropy loss arising from backbone immobilization (ΔS_{bb}) tabulated in ref *41* should possibly be reduced for all RG residues, since packing at the binding site is looser than in an average protein, and the thermal motions of the peptide are not significantly restricted (Figure 7). The conclusion is that there must be a source of negative entropy counterbalancing the positive entropy of surface dehydration. Similarly to ΔC_p and ΔH, water entrapment is not the likely contributor.

Tightening of PDZ2 Possibly Contributes to the Energetic Profile of the PDZ2/RG Complex. Altogether, the comparison between experiment and prediction could be interpreted with the presence of a molecular process manifested by an exothermic heat effect, which is linked to an entropy loss and a negative heat capacity change. The energetic expenditure of the overall structural rearrangement of the free components within the complex, where PDZ2 and RG are in the binding-competent conformations, appears to be small (compare the results of the rb and conf models). Qualitatively, the energetic picture might possibly indicate structural tightening of the PDZ2 domain as the consequence of RG binding. Closer packing intensifies enthalpic interactions, but a more rigid and less fluctuating complex is disfavored

---

[2] Usually, water molecules are involved in the coordination of the C-terminal carboxylate of the incoming peptide within the carboxylate binding loop. In our MD simulations the carboxylate of Val 0 is bonded to the side chain of Asn 16 via a water bridge during 9 ns of the total 15 ns of simulation. For an additional 2 ns a water-mediated bridge is seen with the backbone amide of Gly 19. Water occasionally bridges also the carbonyl group of Ala −1 with the guanidine group of Arg 79.

entropically. The associated heat capacity decrement is not very large since the proposed subtle increase in packing affects mostly buried residues and does not include changes in hydration. A quantitative treatment of the data is not possible to evaluate the associated free energy change, but with the numbers taken at face value from Table 4, it appears that enthalpic stabilization dominates over entropic destabilization at 25 °C. The deduced energetic signature might represent thermodynamic manifestation of the more general concept that "stiffening" of a receptor and cooperativity in ligand binding are two faces of the same phenomenon, as proposed by Cooper and Dryden (*83*) and more recently discussed in detail by Williams et al. in the light of new experimental findings (ref *84* and references cited therein). The experimental evidence for ligand-induced structural tightening of the PDZ2 domain is circumstantial at present. Visual inspection of Figure 7 reveals a uniform sequential distribution of regions with reduced dynamics. Peptide binding to PDZ2 causes depression of the local unfolding/breathing rate around helix $\alpha1$, far away from the peptide binding site (*17*). Significant change in PDZ2 dynamics upon binding was demonstrated on the basis of spin relaxation measurements (*18*). Coupling between binding and changes in the collective dynamics of the domain was also identified by normal mode analysis (*85*). Long-range residue−residue coupling was detected very recently by pump−probe MD simulations (*86*). Fuentes et al. found that the residues which are dynamically affected are clustered in two regions, both distal from the binding site (*18*). The connectivity pattern of residues undergoing changes in dynamics closely resembles the pattern of statistical couplings that has been identified in the PDZ domain family from multiple sequence analysis (*87*). Statistically significant, correlated, and evolutionarily conserved couplings between residues occupying remote protein regions are thought to reflect a mechanism of allosteric regulation and/or cooperativity without large-scale conformational rearrangement. In physical terms, however, the relevant indicator of long-range interactions is the existence of energetic, i.e., thermodynamic couplings. At least in one case of the PDZ family, many statistically coupled positions are in fact thermodynamically coupled (*87*). On the basis of the qualitative agreement between residue dynamics and thermodynamic residue−residue coupling, Fuentes et al. plausibly suggest modulation and diversification of the functional role of PDZ domains through long-range communication between the peptide binding site and "secondary, distal surfaces". Alternatively, localized (or global) response of the PDZ domain may represent a general mechanism to enhance or weaken binding (*84*). The enthalpy−entropy balance of such long-range energetic coupling is not known. It should be noted that statistical couplings describe a protein fold as a whole and must not hold for individual family members; i.e., statistical couplings (a family property) must not translate equally in thermodynamic coupling (a family member property). Moreover, in view of the generally low sequence homology in the PDZ family, the strength of thermodynamic coupling is likely to vary from one family member to another. It is tempting to speculate that the different energetic signature of PDZ−peptide complexes, as discussed above, originates in part from differences in the energetic balance of subtle structural response of PDZ domains to peptide binding.

ACKNOWLEDGMENT

We thank V. Sathya Devi for the preparation of peptide samples and Serge Chesnov for mass spectrometry measurements and amino acid analysis.

REFERENCES

1. Sheng, M., and Sala, C. (2001) PDZ domains and the organization of supramolecular complexes, *Annu. Rev. Neurosci. 24*, 1−29.
2. Bezprozvanny, I., and Maximov, A. (2001) PDZ domains: More than just a glue, *Proc. Natl. Acad. Sci. U.S.A. 98*, 787−789.
3. Hung, A. Y., and Sheng, M. (2002) PDZ domains: Structural modules for protein complex assembly, *J. Biol. Chem. 277*, 5699−5702.
4. Montell, C. (2000) A PDZ protein ushers in new links, *Nat.e Genet. 26*, 6−7.
5. Nourry, C., Grant, S. G. N., and Borg, J.-P. (2003) PDZ domain proteins: Plug and play!, *Sci. STKE 2003*, re7.
6. Sherman, D. L., Fabrizi, C., Gillespie, C. S., and Brophy, P. J. (2001) Specific disruption of a Schwann cell dystrophin-related protein complex in a demyelinating neuropathy, *Neuron 30*, 677−687.
7. Verpy, E., Leibovici, M., Zwaenepoel, I., Liu, X. Z., Gal, A., Salem, N., Mansour, A., Blanchard, S., Kobayashi, I., Keats, B. J. B., Slim, R., and Petit, C. (2000) A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C, *Nat. Genet. 26*, 51−55.
8. Bezprozvanny, I., and Maximov, A. (2001) Classification of PDZ domains, *FEBS Lett. 509*, 457−462.
9. Birrane, G., Chung, J., and Ladias, J. A. A. (2003) Novel mode of ligand recognition by the erbin PDZ domain, *J. Biol. Chem. 278*, 1399−1402.
10. Kozlov, G., Banville, D., Gehring, K., and Ekiel, I. (2002) Solution structure of the PDZ2 domain from cytosolic human phosphatase hPTP1E complexed with a peptide reveals contribution of the beta 2-beta 3 loop to PDZ domain-ligand interactions, *J. Mol. Biol. 320*, 813−820.
11. Kozlov, G., Gehring, K., and Ekiel, I. (2000) Solution structure of the PDZ2 domain from human phosphatase hPTP1E and its interactions with C-terminal peptides from the Fas receptors, *Biochemistry 39*, 2572−2580.
12. Skelton, N. J., Koehler, M. F. T., Zobel, K., Wong, W. L., Yeh, S., Pisabarro, M. T., Yin, J. P., Lasky, L. A., and Sidhu, S. S. (2003) Origins of PDZ domain ligand specificity: structure determinationand mutagenesis of the Erbin PDZ domain, *J. Biol. Chem. 278*, 7645−7654.
13. Tochio, H., Hung, F., Li, M., Bredt, D. S., and Zhang, M. J. (2000) Solution structure and backbone dynamics of the second PDZ domain of postsynaptic density-95, *J. Mol. Biol. 297*, 830−830.
14. Erdmann, K. S., Kuhlmann, J., Lessmann, V., Herrmann, L., Eulenburg, V., Muller, O., and Heumann, R. (2000) The adenomatous polyposis coli-protein (APC) interacts with the protein tyrosine phosphatase PTP-BL via an alternatively spliced PDZ domain, *Oncogene 19*, 3894−3901.
15. Kim, E., DeMarco, S. J., Marfatia, S. M., Chishti, A. H., Sheng, M., and Strehler, E. E. (1998) Plasma membrane Ca^{2+} ATPase isoform 4b binds to membrane-associated guanylate kinase (MAGUK) proteins via their PDZ (PSD-95/Dlg/ZO-1) domains, *J. Biol. Chem. 273*, 1591−1595.
16. Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., Crompton, A., Chan, A. C., Anderson, J. M., and Cantley, L. C. (1997) Recognition of unique carboxyl-terminal motifs by distinct PDZ domains, *Science 275*, 73−77.
17. Ekiel, I., Banville, D., Shen, S. H., and Gehring, K. (1998) Effect of peptide binding on amide proton exchange rates in the PDZ2 domain from human phosphatase hPTP1E, *Biochem. Cell. Biol. 76*, 334−340.
18. Fuentes, E. J., Der, C. J., and Lee, A. L. (2004) Ligand-dependent dynamics and intramolecular signaling in a PDZ domain, *J. Mol. Biol. 335*, 1105−1115.
19. Kang, B. S., Cooper, D. R., Jelen, F., Devedjiev, Y., Derewenda, U., Dauter, Z., Otlewski, J., and Derewenda, Z. S. (2003) PDZ tandem of human syntenin: Crystal structure and functional properties, *Structure 11*, 459−468.
20. Niethammer, M., Valtschanoff, J. G., Kapoor, T. M., Allison, D. W., Weinberg, R. J., Craig, A. M., and Sheng, M. (1998) CRIPT,

a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90, *Neuron 20*, 693−707.

21. Lu, J., Li, H., Wang, Y., Sudhof, T. C., and Rizo, J. (2005) Solution structure of the RIM1[alpha] PDZ domain in complex with an ELKS1b C-terminal peptide, *J. Mol. Biol. 352*, 455−466.
22. Freiss, G., and Vignon, F. (2004) Protein tyrosine phosphatases and breast cancer, *CRC Oncol. Hematol. 52*, 9−17.
23. Palmer, A., Zimmer, M., Erdmann, K. S., Eulenburg, V., Porthin, A., Heumann, R., Deutsch, U., and Klein, R. (2002) EphrinB phosphorylation and reverse signaling: Regulation by Src kinases and PTP-BL phosphatase, *Mol. Cell 9*, 725−737.
24. Sato, T., Irie, S., Kitada, S., and Reed, J. C. (1995) FAP-1: a protein tyrosine phosphatase that associates with Fas, *Science 268*, 411−415.
25. Murthy, K. K., Clark, K., Fortin, Y., Shen, S.-H., and Banville, D. (1999) ZRP-1, a zyxin-related protein, interacts with the second PDZ domain of the cytosolic protein tyrosine phosphatase hPTP1E, *J. Biol. Chem. 274*, 20679−20687.
26. Gao, X., Satoh, T., Liao, Y., Song, C., Hu, C.-D., Kariya, K.-i., and Kataoka, T. (2001) Identification and characterization of RA-GEF-2, a Rap guanine nucleotide exchange factor that serves as a downstream target of M-Ras, *J. Biol. Chem. 276*, 42219−42225.
27. Plotnikov, V. V., Brandts, J. M., Lin, L. N., and Brandts, J. F. (1997) A new ultrasensitive scanning calorimeter, *Anal. Biochem. 250*, 237−244.
28. Freire, E. (1995) Thermal denaturation methods in the study of protein folding, *Methods Enzymol. 259*, 144−169.
29. Freire, E., and Biltonen, R. L. (1978) Statistical mechanical deconvolution of thermal transitions in macromolecules. I. Theory and application to homogeneous systems, *Biopolymers 17*, 463−479.
30. Privalov, P. L., and Potekhin, S. A. (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins, *Methods Enzymol. 131*, 4−51.
31. Privalov, P. L., and Makhatadze, G. I. (1990) Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: Protein unfolding effects, *J. Mol. Biol. 213*, 385−391.
32. Milev, S., Gorfe, A. A., Karshikoff, A., Clubb, R. T., Bosshard, H. R., and Jelesarov, I. (2003) Energetics of sequence-specific protein-DNA association: Binding of integrase Tn916 to its target DNA, *Biochemistry 42*, 34813491.
33. Brandts, J. F., and Lin, L. N. (1990) Study of strong to ultratight protein interactions using differential scanning calorimetry, *Biochemistry 29*, 6927−6940.
34. Bashford, D. (2004) Macroscopic electrostatic models for protonation states in proteins, *Front. Biosci. 9*, 1082−1099.
35. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis, *J. Mol. Model. 7*, 306−317.
36. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926−935.
37. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem. 18*, 1463−1472.
38. Miyamoto, S., and Kollman, P. A. (1992) Settle - an analytical version of the shake and rattle algorithm for rigid water models, *J. Comput. Chem. 13*, 952−962.
39. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald—an N.Log(N) method for Ewald sums in large systems, *J. Chem. Phys. 98*, 10089−10092.
40. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular-dynamics with coupling to an external bath, *J. Chem. Phys. 81*, 3684−3690.
41. Baker, B. M., and Murphy, K. P. (1998) Prediction of binding energetics from structure using empirical parameterization, *Methods Enzymol. 295*, 294−315.
42. Luque, I., and Freire, E. (1998) Structure-based prediction of binding affinities and molecular design of peptide ligands, *Methods Enzymol. 295*, 100−127.
43. Hubbard, S. J., and Thornton, J. M. (1993) NACCESS, Computer Program, Department of Biochemistry and Molecular Biology, University College, London.
44. Yu, Y. B., Privalov, P. L., and Hodges, R. S. (2001) Contribution of translational and rotational motions to molecular association in aqueous solution, *Biophys. J. 81*, 1632−1642.
45. Antonsson, P., Kammerer, R. A., Schulthess, T., Hanisch, G., and Engel, J. (1995) Stabilization of the alpha-helical coiled-coil domain in laminin by C-terminal disulfide bonds, *J. Mol. Biol. 250*, 74−79.
46. Bardi, J. S., Luque, I., and Freire, E. (1997) Structure-based thermodynamic analysis of HIV-1 protease inhibitors, *Biochemistry 36*, 6588−6596.
47. Gomez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) The heat capacity of proteins, *Proteins: Struct., Funct., Genet. 22*, 404−412.
48. Gianni, S., Calosci, N., Aelen, J. M. A., Vuister, G. W., Brunori, M., and Travaglini-Allocatelli, C. (2005) Kinetic folding mechanism of PDZ2 from PTP-BL, *Protein Eng., Des. Sci. 18*, 389−395.
49. Makhatadze, G. I., and Privalov, P. L. (1995) Energetics of protein structure, *Adv. Protein Chem. 47*, 307−425.
50. Robertson, A. D., and Murphy, K. P. (1997) Protein structure and the energetics of protein stability, *Chem. Rev. 97*, 1251−1267.
51. Haynie, D. T., and Freire, E. (1994) Estimation of the folding/unfolding energetics of marginally stable proteins using differential scanning calorimetry, *Anal. Biochem. 216*, 33−41.
52. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m-values and heat-capacity changes—relation to changes in accessible surface-areas of protein unfolding, *Protein Sci. 4*, 2138−2148.
53. Gianni, S., Engstrom, A., Larsson, M., Calosci, N., Malatesta, F., Eklund, L., Ngang, C. C., Travaglini-Allocatelli, C., and Jemth, P. (2005) The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism, *J. Biol. Chem. 280*, 34805−34812.
54. Saro, D., Klosi, E., Paredes, A., and Spaller, M. R. (2004) Thermodynamic analysis of a hydrophobic binding site: Probing the PDZ domain with nonproteinogenic peptide ligands, *Org. Lett. 6*, 3429−3432.
55. Marfatia, S. M., MoraisCabral, J. H., Kim, A. C., Byron, O., and Chishti, A. H. (1997) The PDZ domain of human erythrocyte p55 mediates its binding to the cytoplasmic carboxyl terminus of glycophorin C—Analysis of the binding interface by in vitro mutagenesis, *J. Biol. Chem. 272*, 24191−24197.
56. Gopalakrishnan, M., Forsten-Williams, K., Cassino, T., Padro, L., Ryan, T., and Tãuber, U. (2005) Ligand rebinding: self-consistent mean-field theory and numerical simulations applied to surface plasmon resonance studies, *Eur. Biophys. J. 34*, 943−958.
57. Nieba, L., Krebber, A., and Pluckthun, A. (1996) Competition BIAcore for measuring true affinities: Large differences from values determined from binding kinetics, *Anal. Biochem. 234*, 155−165.
58. Carra, J. H., and Privalov, P. L. (1997) Energetics of folding and DNA binding of the MAT alpha homeodomain, *Biochemistry 36*, 526−535.
59. Celej, M. S., Dassie, S. A., Gonzalez, M., Bianconi, M. L., and Fidelio, G. D. (2006) Differential scanning calorimetry as a tool to estimate binding parameters in multiligand binding proteins, *Anal. Biochem. 350*, 277−284.
60. Mitkevich, V. A., Schulga, A. A., Ermolyuk, Y. S., Lobachov, V. M., Chekhov, V. O., Yakovlev, G. I., Hartley, R. W., Nick, Pace, C., Kirpichnikov, M. P., and Makarov, A. A. (2003) Thermodynamics of denaturation of complexes of barnase and binase with barstar, *Biophys. Chem. 105*, 383−390.
61. Plotnikov, V., Rochalski, A., Brandts, M., Brandts, J. F., Williston, S., Frasca, V., and Lin, L. N. (2002) An autosampling differential scanning salorimeter instrument for studying molecular interactions, *Assay Drug Dev. Technol. 1* (Part 1), 83−90.
62. Waldron, T. T., and Murphy, K. P. (2003) Stabilization of proteins by ligand binding: Application to drug screening and determination of unfolding energetics, *Biochemistry 42*, 5058−5064.
63. Rupasinghe, C. N., and Spaller, M. R. (2006) The interplay between structure-based design and combinatorial chemistry, *Curr. Opin. Struct. Biol. 10*, 1−6.
64. Wiedemann, U., Boisguerin, P., Leben, R., Leitner, D., Krause, G., Moelling, K., Volkmer-Engert, R., and Oschkinat, H. (2004) Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides, *J. Mol. Biol. 343*, 703−718.
65. Ruben, A. J., Kiso, Y., and Freire, E. (2006) Overcoming roadblocks in lead optimization: A thermodynamic perspective, *Chem. Biol. Drug Des. 67*, 2−4.

66. Cummings, M. D., Hart, T. N., and Read, R. J. (1995) Atomic solvation parameters in the analysis of protein-protein docking results, *Protein Sci. 4*, 2087−2099.
67. Horton, N., and Lewis, M. (1992) Calculation of the free-energy of association for protein complexes, *Protein Sci. 1*, 169−181.
68. Krystek, S., Stouch, T., and Novotny, J. (1993) Affinity and specificity of serine endopeptidase protein inhibitor interactions—Empirical free-energy calculations based on X-ray crystallographic structures, *J. Mol. Biol. 234*, 661−679.
69. Luque, I., and Freire, E. (2002) Structural parameterization of the binding enthalpy of small ligands, *Proteins: Struct., Funct., Genet. 49*, 181−190.
70. Williams, D. H., Cox, J. P. L., Doig, A. J., Gardner, M., Gerhard, U., Kaye, P. T., Lal, A. R., Nicholls, I. A., Salter, C. J., and Mitchell, R. C. (1991) Toward the semiquantitative estimation of binding constants—Guides for peptide peptide binding in aqueous solution, *J. Am. Chem. Soc. 113*, 7020−7030.
71. Walma, T., Spronk, C. A. E. M., Tessari, M., Aelen, J., Schepens, J., Hendriks, W., and Vuister, G. W. (2002) Structure, dynamics and binding characteristics of the second PDZ domain of PTP-BL, *J. Mol. Biol. 316*, 1101−1110.
72. Baker, B. M., and Murphy, K. P. (1997) Dissecting the energetics of a protein-protein interaction: the binding of ovomucoid third domain to elastase, *J. Mol. Biol. 268*, 557−569.
73. Lavigne, P., Bagu, J. R., Boyko, R., Willard, L., Holmes, C. F., and Sykes, B. D. (2000) Structure-based thermodynamic analysis of the dissociation of protein phosphatase-1 catalytic subunit and microcystin-LR docked complexes, *Protein Sci. 9*, 252−264.
74. Sturtevant, J. M. (1977) Heat-capacity and entropy changes in processes involving proteins, *Proc. Natl. Acad. Sci. U.S.A. 74*, 2236−2240.
75. Eftink, M., and Biltonen, R. L. (1980) Thermodynamics of interaction biological systems., in *Biological Microcalorimetry* (Beezer, A. E., Ed.) pp 343−412, Academic Press, San Diego, CA.
76. Cooper, A., Johnson, C. M., Lakey, J. H., and Nollmann, M. (2001) Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions, *Biophys. Chem. 93*, 215−230.
77. Privalov, G. P., and Privalov, P. L. (2000) Problems and prospects in microcalorimetry of biological macromolecules, *Methods Enzymol. 323*, 31−62.
78. Cooper, A. (2005) Heat capacity effects in protein folding and ligand binding: A re-evaluation of the role of water in biomolecular thermodynamics, *Biophys. Chem. 115*, 89−97.
79. Morton, C. J., and Ladbury, J. E. (1996) Water mediated protein-DNA interactions: The relationship of thermodynamics to structural detail, *Protein Sci. 5*, 2115−2118.
80. Connelly, P. R. (1997) *The Cost of Releasing Site-Specific, Bound Water Molecules from Protein: Towards a Quantitative Guide for Structure-Based Drug Design*, Springer, Berlin.
81. Hilser, V. J., Gomez, J., and Freire, E. (1996) Enthalpy change in protein folding and binding: Refinement of parameters for structure-based calculations, *Proteins: Struct., Funct., Genet. 26*, 123−133.
82. Holdgate, G. A., Tunnicliffe, A., Ward, W. H. J., Weston, S. A., Rosenbrock, G., Barth, P. T., Taylor, I. W. F., Pauptit, R. A., and Timms, D. (1997) The entropic penalty of ordered water accounts for weaker binding of the antibiotic novobiocin to a resistant mutant of DNA gyrase: A thermodynamic and crystallographic study, *Biochemistry 36*, 9663−9673.
83. Cooper, A., and Dryden, D. T. F. (1984) Allostery without conformational change. A plausible model, *Eur. Biophys. J. 11*, 103−109.
84. Williams, D. H., Stephens, E., O'Brien, D. P., and Zhou, M. (2004) Understanding noncovalent interactions: Ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes, *Angew. Chem., Int. Ed. 43*, 6596−6616.
85. De Los Rios, P., Cecconi, F., Pretre, A., Dietler, G., Michielin, O., Piazza, F., and Juanico, B. (2005) Functional dynamics of PDZ binding domains: A normal-mode analysis, *Biophys. J. 89*, 14−21.
86. Sharp, K., and Skinner, J. J. (2006) Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling, *Proteins: Struct., Funct., Bioinf. 65*, 347−361.
87. Lockless, S. W., and Ranganathan, R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families, *Science 286*, 295−299.

BI061869I

5.3 THERMODYNAMICS OF B-HLH-LZ PROTEIN BINDING TO DNA: THE ENERGETIC IMPORTANCE OF PROTEIN-DNA CONTACTS IN SITE-SPECIFIC E-BOX RECOGNITION BY THE COMPLETE GENE PRODUCT OF THE MAX P21 TRANSCRIPTION FACTOR

Laura Meier-Andrejszki and Saša Bjelić and Jean-François Naud and Pierre Lavigne and Ilian Jelesarov

Article published in Biochemistry 46: 12427 - 12440

To the following article I contributed analytical ultracentrifugation experiments aimed at characterisation of the oligomeric state of the Max p21 protein and the Max p21/E-box DNA complex. Further, I actively participated in the construction of mutants, in the design analysis of the experiments.

Thermodynamics of b-HLH-LZ Protein Binding to DNA: The Energetic Importance of Protein−DNA Contacts in Site-Specific E-Box Recognition by the Complete Gene Product of the Max p21 Transcription Factor[†]

Laura Meier-Andrejszki,[‡] Saša Bjelić,[‡] Jean-François Naud,[§] Pierre Lavigne,[§] and Ilian Jelesarov*,[‡]

Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, and Département de Pharmacologie, Institut de Pharmacologie de Sherbrooke, Université de Sherbrooke, Canada J1H 5N4

Received June 4, 2007; Revised Manuscript Received August 3, 2007

ABSTRACT: The Myc/Mad/Max network of dimeric basic region−helix−loop−helix−leucine zipper (b-HLH-LZ) transcription factors bind to enhancer box sequences (E-box) in the promotors of a large set of genes that control cell metabolism, proliferation, and differentiation. Max (Myc-associated factor X) is the obligate heterodimerization partner of Myc and Mad proteins. On the other hand, Max is the only member of the family capable of forming a stable homodimer. As part of the transcriptional regulation mechanism, Myc/Max and Mad/Max heterodimers and Max homodimers are thought to compete for binding to the E-box target sequences. E-box recognition is structurally supported by the b-HLH-LZ structural motif, which also promotes dimerization. However, the actual dimerization and heterodimerization constants of the complete gene products and their affinities for E-box sequences are not known. Also, the detailed thermodynamic characterization of DNA binding by these transcription factors has not been done yet. Such knowledge is necessary for complete understanding of the transcriptional regulation carried out by the Myc/Mad/Max network. Here, we report the first in-depth thermodynamic characterization of the stability and specific DNA binding of a full length gene product of the Myc/Mad/Max family, namely, Max protein isoform p21 (Max p21). Using calorimetric methods (DSC and ITC) we have determined the dimerization constant of Max p21 in the low micromolar range, and the Max p21/E-box complex dissociation constant in the low nanomolar range at 37 °C. The association is driven by a large exothermic effect, which is partly compensated by entropic factors. The energetic contribution to binding affinity of seven highly conserved residues that contact the DNA was probed by X-to-Ala mutagenesis. The results demonstrate that high binding affinity critically relies on the side chain of Arg 26. Furthermore, the mutational analysis points to the important role of the persistent helical turn that comprises this residue at the junction of the basic region and helix H1. Altogether, the study supports the idea that Max p21 can bind E-box sequences in vivo and likely participates directly in the regulation of transcription as homodimer.

The Max[1] protein (Myc associated factor X) belongs to the Myc/Mad/Max network of transcription factors controlling cell metabolism, proliferation, and differentiation (*1*). The network includes, next to Max, members of the Myc and Mad families, the Mad-related protein Mnt, and Mga. Max appears to have a central role in concerting the action within the network, since it is the obligate heterodimerization partner of Myc, Mad, and Mnt, thereby promoting recognition of the enhancer box (E-box) sequences in the promoters of a large set of genes, and triggering various activities of the network (*2*). Myc/Max heterodimers activate genes involved in cell growth and proliferation by the recruitment of coactivators of transcription that possess HAT activity. Myc deregulation and overexpression is associated with apoptosis, genetic instability, and malignant transformations, such as small cell lung cancer, neuroblastoma, and glioblastoma (*3*). The Mad/Max heterodimer antagonizes most of Myc/Max transcriptional activities by recruiting corepressors with HDAC activity (*4*). The Mnt/Max heterodimer also has a tumor suppressor activity (ref (*5*) and citations therein). While Myc, Mad, and Mnt proteins have short half-lives and their expression levels are strictly controlled by diverse signal transduction cascades like the MAP-kinase or TGF-β pathway, Max itself has a significantly longer half-life and is stably and ubiquitously expressed. Max is the only member of the family which is capable of forming sufficiently stable homodimers (*in vitro*). However, *in vitro* studies with truncated constructs demonstrate that the $(\mathrm{Max})_2$ homodimer is less stable than the Myc/Max and Mad/Max heterodimers (*6*). Being transcriptionally inert and lacking the domains

---

[†] This work was supported by the Swiss National Center of Competence in Research "Structural Biology" and in part by Grant 31-100197/1 from the Swiss National Science Foundation and the Canadian Institutes for Health Research.

* To whom correspondence should be addressed. Phone: ++41 44 635 5547. Fax: ++41 44 635 6805. E-mail: iljel@bioc.unizh.ch.

[‡] Biochemisches Institut der Universität Zürich.

[§] Université de Sherbrooke.

[1] Abbreviations: b-HLH-LZ; basic−helix−loop−helix−leucine zipper; DSC, differential scanning calorimetry; E-box, enhancer box; HAT, histone acetyl transferase; HDAC, histone deacetylase; ITC, isothermal titration calorimetry; Mad, mitotic absence deficient; Max, Myc-associated factor X; MRE, mean residue ellipticity; Myc, myelocytoma-associated protein; ΔG, free energy change; ΔH, enthalpy change; ΔS, entropy change; ΔC_p, heat capacity change.

10.1021/bi701081q CCC: $37.00 PAGE EST: 13.5

necessary for recruiting coactivator or corepressor protein complexes, Max is hypothesized to be able to modulate cell growth by competing for E-box with Myc/Max, Mad/Max, or Mnt/Max. It is still unknown whether the competition proceeds at the level of heterodimer-*versus*-homodimer E-box binding affinity discrimination, or else $(Max)_2$ binding to DNA reduces the concentration of Max monomers available for heterodimerization with other family members.

All members of the Myc/Max/Mad network contain a basic region−helix−loop−helix−leucine zipper (b-HLH-LZ) structural motif, which is responsible for DNA binding and dimerization, flanked by sequences that are involved in recruiting of effector partner proteins. The crystal structures of c-Myc/Max, Mad1/Max, and $(Max)_2$ proteins bound to E-box have revealed a very similar structural organization of the b-HLH-LZ domains in the context of either homodimeric $(Max)_2$ or of the heterodimers (*7*−*9*). The HLH domain folds as a parallel four-helix bundle. A parallel, two-stranded coiled coil is formed by the adjacent LZ domains. The basic region domains form α-helices naturally emerging from the end of helix H1 of HLH and diverging to grip the duplex by fitting into the major groove. It has been long suggested that the basic region segments undergo a conformational transition from a quasi-random coil in the free protein to an α-helix in the protein−DNA complex (*10*, *11*). Very recently, the structure of the unbound $(Max)_2$ b-HLH-LZ homodimer was determined by NMR, for the first time allowing insights into the structural rearrangements accompanying DNA recognition (*12*). Indeed, the first 14 residues of the basic region are largely unfolded in the absence of DNA. Interestingly, however, the last four residues of the basic region form a highly populated helical turn, just adjacent to helix H1 of HLH. Since highly conserved basic residues contacting the E-box site are located in this region, it was plausibly suggested that prefolded α-helical conformers might speed up specific binding. The loop region possesses high flexibility in the free state. Little difference was observed in the conformation of the LZ coiled coil domain in the free and in the bound state.

Specific E-box binding by Myc/Mad/Max family members is mediated by conserved side chains from the basic region. Structural studies have identified a conserved and quasi-symmetric recognition pattern within the 5′-CACGTG-3′ target, including four specific hydrogen bonds to DNA bases and a number of nonspecific electrostatic contacts with the phosphate backbone. In addition, residues from the loop region closely approach the duplex and are seemingly involved in binding. There is some controversy about the number of protein−DNA backbone contacts: Six such interactions have been identified in the structure of the $(Max)_2$−DNA complex (*7*). While additional contacts, both from the basic region and from the loop regions, are made by the Myc protein in the Myc/Max−DNA complex (*9*), there is no structural evidence that the regions flanking the b-HLH-LZ domain are involved in E-box recognition.

Although the structural basis of E-box recognition by b-HLH-LZ proteins of the Myc/Mad/Max family is now firmly established, much less is known about the energetics of binding and the relative affinities of different dimers for the cognate DNA sequence. So far, only the binding affinity of short constructs encompassing the b-HLH-LZ core domain has been determined (*13*−*15*). In the present study we use calorimetry to characterize thermodynamically the site-specific binding of the complete Max p21 isoform to a 21 base pair DNA duplex containing the E-box sequence. By examining the thermal stability of Max p21 and Max p21/DNA complex, we provide for the first time a reliable estimate of the dissociation constant at the physiologically relevant temperature of 37 °C. X-to-Ala mutations reveal the energetic importance of specific and nonspecific protein−DNA contacts. Direct binding experiments demonstrate that stabilization of the leucine zipper domain increases the DNA-binding affinity.

EXPERIMENTAL PROCEDURES

Buffer. All biophysical experiments were performed in a standard buffer containing 100 mM sodium phosphate, pH 6.8. The buffer was supplemented with 100−600 mM KCl. The pH of samples containing urea was adjusted after addition of the denaturant. The urea concentration was determined by measuring the refractive index. All chemicals were of analytical grade and were used without further purification.

Molecular Biology, Protein Expression, and Purification. The construction of the pET3a expression plasmid and expression of Max p21, Max p21VL, and Max_{SS} has been described (*16*, *17*). Site directed mutagenesis was carried out using commercially available kits from QIAGEN following the protocol recommended by the manufacturer. The primers designed for mutagenesis PCR are available on request. After expression, the cells were frozen and stored at −20 °C. Purification was done as described before (*16*, *17*). In an additional purification step, the combined fractions after ion exchange chromatography were dialyzed against water overnight, loaded on a semipreparative C8-HPLC column, and eluted with binary acetonitrile/water gradients (3% to 60%) containing 0.1% and 0.085% trifluoroacetic acid. The material after HPLC was lyophilized. The protein refolded from the lyophilized state was fully native, as judged by CD spectroscopy, and displayed DNA binding affinity indistinguishable from that measured with natively purified non-lyophilized material. The purity was verified by analytical HPLC with detection at 220 nm, and the mass was verified by mass spectrometry. The concentration was measured by UV spectroscopy using $\epsilon_{280} = 5960\ M^{-1}\ cm^{-1}$ calculated from the amino acid sequence.

DNA Synthesis and Annealing. The 21 bp DNA single strand oligonucletides bearing the E-box target site (see Figure 1) were custom-synthesized (Metabion GmbH). After HPLC purification, they were dissolved in water and the concentration was determined by UV after complete digestion with phosphodiesterase overnight at 37 °C. The annealing was done by mixing equimolar amounts of both strands in 100 mM potassium chloride. The sample was incubated at 90 °C for 5 min and then slowly cooled down over 4 h. Concentration of the double stranded E-box was determined by UV spectroscopy after digestion with phosphodiesterase, and the annealing was confirmed by agarose gel electrophoresis.

Circular Dichroism (CD) Spectroscopy. Circular dichroism spectroscopy was perfomed on a Jasco J-715 spectropolarimeter equipped with a computer-controlled water thermostat using jacketed quartz cells of 0.2, 0.5, or 1 cm optical path

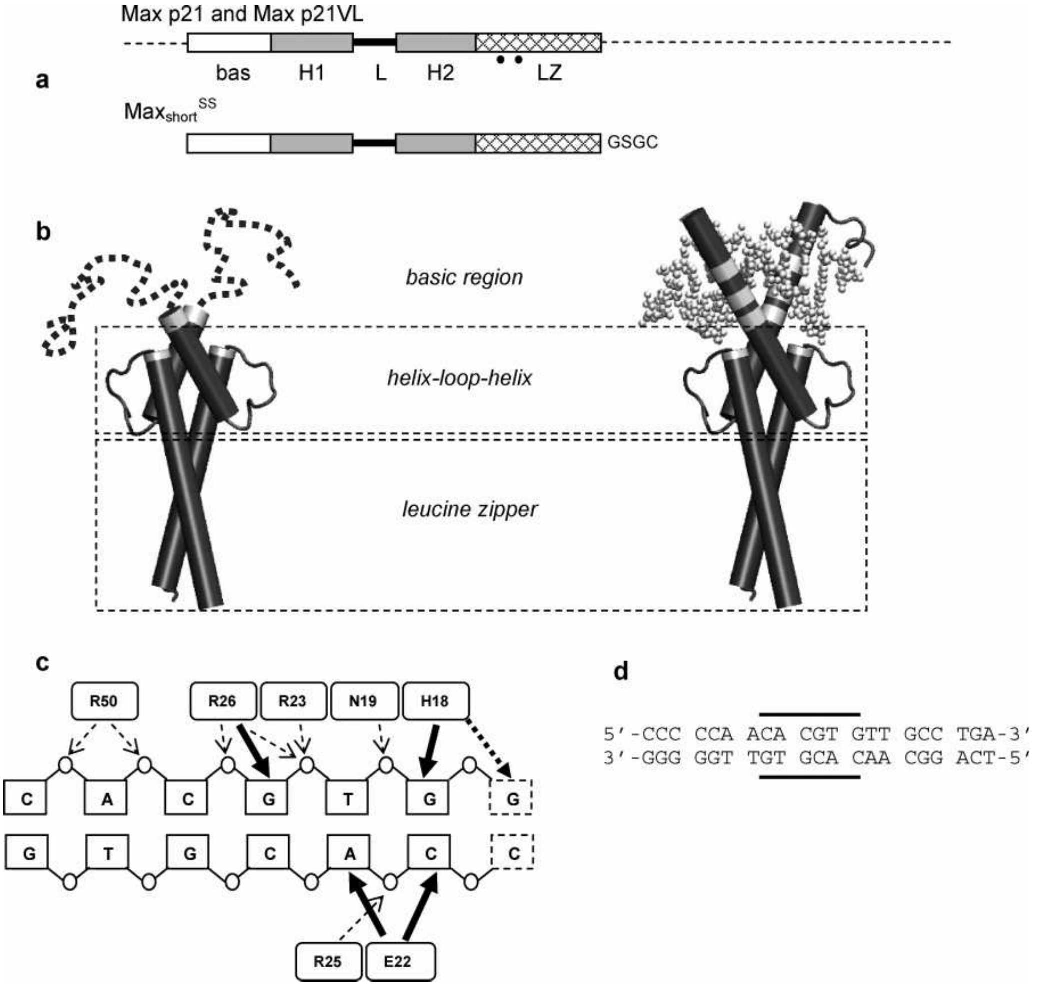


FIGURE 1: Structure of Max p21 and protein−DNA interactions in the Max p21-E-box complex. (a) Sequential organization. White bar, basic region (bas). Gray bars, α-helices (H1 and H2). Hatched bar, leucine zipper (LZ). Thick line, loop (L). The dashed lines are the unstructured N- and C-terminal segments. The length of the schematic elements is proportional to the corresponding sequence length. The positions of the N69V and H72L mutations in Max p21VL are indicated with dots. In the lower scheme, the structure of the short constructs is shown. Max$_{short}$ lacks the unstructured segments outside the b-HLH-LZ core domain. The GSGC C-terminal extension facilitates formation of the disulfide linked Max$_{short}^{SS}$ protein. (b) E-box binding does not perturb the structure of HLH and LZ, but the basic region folds from a predominantly coil conformation to a regular α-helix fitting into the DNA major groove. In the cartoon representation residues contacting DNA are shown in white. Note that the very C-terminus of the basic region containing the DNA-contacting Arg 25 and Arg 26 is part of an α-helical turn also in the free protein. (c) Protein−DNA interactions within the E-box target sequence. The protein−DNA contacts formed by one of the protein monomers are shown as bold arrows (base-specific hydrogen bonds) and thin arrows (contacts to the phosphate backbone. (d) Sequence of the DNA duplex used in this study. The E-box is demarcated.

length. Spectra were recorded at a scanning speed of 5 nm min^{-1} and 0.2 nm intervals. Three scans were averaged to obtain the final spectrum. Isothermal urea unfolding experiments were performed at 5, 20, and 25 °C. The samples were incubated overnight and the ellipticity at 222 nm was sampled for 3 min after thermal equilibration. The data were analyzed following the linear extrapolation model (LEM) as detailed elsewhere (*18*). The concentration dependence of ellipticity at 222 nm was studied with 1−100 μM (monomer equivalents) protein solutions. Thermal denaturation was performed by monitoring the change in ellipticiy at 222 nm during continuous heating at 1 deg min^{-1} between 5 and 75−90 °C. Reversibility was checked from the recovery of the CD signal after cooling, and was at least 90% in all cases.

Isothermal Titration Calorimetry. ITC experiments were performed on a VP-ITC instrument (MicroCal Inc., Northampton, MA). The calorimeter was calibrated according to the manufacturer's instruction. Samples of protein and DNA were prepared in, and thoroughly dialyzed against, the same batch of buffer to minimize artifacts due to minor differences in buffer composition. The concentration was determined after dialysis. The sample cell (1.4 mL) was loaded with 8−20 μM (dimer equivalents) protein; DNA duplex concentration in the syringe was 90−300 μM. A titration experiment typically consisted of 25−30 injections, each of 8 or 10 μL volume and 10 or 12 s duration, with a 5 min interval between additions. Stirring rate was 300 rpm. Raw data were integrated, corrected for nonspecific heats, normalized for concentration, and analyzed according to a 1:1 binding model assuming a single set of identical binding sites.

Differental Scanning Calorimetry. DSC experiments were performed on a VP-DSC calorimeter (MicroCal Inc.) equipped with twin coin-shaped cells of 0.52 mL volume. Details on the instrument's performance are given elsewhere (*19*). The heating rate was 1 °C min^{-1}. Samples containing protein and DNA (in isolation or as a mixture) were dialyzed for 18−24 h against the same batch of buffer used to establish the instrumental buffer−buffer baseline. Reversibility was checked by 2−3 cycles of heating and cooling. The raw experimental data were corrected for the instrumental baseline and transformed to partial molar or partial specific heat capacity using partial specific volumes calculated from the amino acid composition. The analysis of heat capacity traces followed the formalism detailed elsewhere (*20*, *21*). Data handling and analysis were carried out using the program CpCalc 2.1 (Applied Thermodynamics), subroutines for Origin provided by MicroCal, and in-house written scripts for NLREG (Phillip H. Sherrod).

RESULTS AND DISCUSSION

A schematic representation of the structural organization of the full gene product of the Max p21 protein and its b-HLH-LZ domain (henceforth referred to as Max p21 and Max$_{short}$, respectively) is shown in Figure 1. The figure indicates also the position of the cysteine residue which was introduced at the C-terminus of the leucine zipper to produce the disulfide-linked variant Max$_{short}^{SS}$ after oxidation, as well as the location of the N68V and H71L mutations in the Max p21VL variant. All proteins were expressed in soluble form in the cytoplasm of *Escherichia coli*. Purified Max p21 exists as a homodimer in the concentration range 15 to 120 μM (monomer equivalents) under the experimental conditions used in this study, as evidenced by equilibrium and velocity analytical ultracentrifugation (see Figure S1 in Supporting Information).[2] We first describe spectroscopic characterization and unfolding experiments aimed at determination of the thermodynamic stability of Max p21. This information is important in order to carry out the binding experiments in the temperature range where Max p21 is in its fully native and dimeric state. ITC and DSC experiments revealing the energetics of Max p21 binding to E-box DNA are presented in the second part. In the third part, we evaluate the energetic role of conserved Max p21 side chains to the DNA binding affinity from binding experiments with Max p21 variants, in which particular protein−DNA contacts were removed by mutation to alanine.

Stability and Unfolding Energetics of Max p21

The N- and C-Terminal Extensions of the Max b-HLH-LZ Domain Are Disordered. The CD spectrum of Max$_{short}$ is typical for a protein with high helical content (Figure 2, trace b). The per-residue molar ellipticity (MRE) of Max p21 is significantly lower (trace a). However if one expresses the CD signal of Max p21 normalized for 86 residues (the length of Max$_{short}$), instead of 150 residues (the length of Max p21), one can appreciate the spectroscopic signature of the N- and C-terminal segments that are not present in Max$_{short}$ (trace c). The normalized MRE$_{222}$ of Max p21 (trace c) is slightly more negative (indicating higher helical content) than MRE$_{222}$ of Max$_{short}$ (trace b). The difference is likely due to the fact that the fraction of unfolded (monomeric and nonhelical) Max$_{short}$ is 5−10% at the selected concentration (100 μM; ref (*17*)), whereas Max p21 is more than 98% dimeric (20 μM; see Figure 5). In contrast, the decrease in ellipticity below 210 nm is pronounced and is accompanied by a blue shift as expected for the existence of unfolded segments. It follows that the ∼70 "extra" residues beyond the b-HLH-LZ core domain of Max p21 are apparently devoid of stable secondary structure elements.

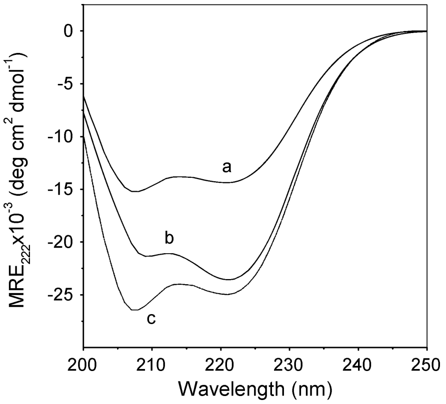


FIGURE 2: Far-UV CD spectra of Max p21 (trace a) and Max$_{short}$ (trace b) at 25 °C in standard buffer. Concentrations were 20 μM Max p21 and 100 μM Max$_{short}$ (monomer equivalents). Trace c represents MRE$_{222}$ of Max p21 recalculated for 86 amino acid residues per monomer (see the text for details).

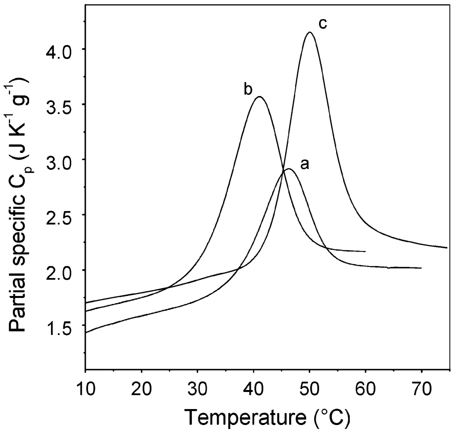


FIGURE 3: Partial specific heat capacity of Max p21 (trace a), Max$_{short}$ (trace b), and Max$_{short}^{SS}$ (trace c). Concentrations were 125 μM Max p21, 280 μM Max$_{short}$ (monomer equivalents), and 100 μM Max$_{short}^{SS}$.

The participation of unstructured polypeptide domains in intra- or intermolecular contacts cannot be excluded *a priori* on the basis of spectroscopic data. Scanning calorimetry experiments were performed to check whether the N- and C-terminal extensions of the b-HLH-LZ domain contribute to the unfolding enthalpy. Max p21 exhibits a significantly lower partial specific enthalpy than Max$_{short}$ (Figure 3). When compared at the same temperature (43 °C, selected in between the corresponding melting temperatures to minimize the errors of extrapolation) these enthalpies amount to 10.5

---

[2] We have not performed AUC experiments with Max$_{short}$ and Max$_{short}^{SS}$. However, the partial molar heat capacity of both proteins between 5 and 25 °C shows no concentration dependence up to ∼300 μM monomer concentration. All DSC traces are well described by a two-state unfolding model. Such behavior is a strong evidence that oligomerization does not take place.

$\pm$ 0.5 J g^{-1} (Max p21) and 17.1 $\pm$ 0.7 10. J g^{-1} (Max$_{\text{short}}$). The ratio is thus 0.61, close to the molar mass ratio (0.59). Furthermore, the temperature slopes of the partial specific heat capacities in the native region are very close to each other, ~13 J K^{-2} g^{-1}. Hence, within the limit of detection there are no enthalpically rich packing interactions involving molecular segments outside the core b-HLH-LZ domain. Altogether, the presented calorimetric results provide sound evidence that the N- and C-terminal domains of Max p21 are unfolded, as previously suggested on the basis of spectroscopic data only (*16*). Interestingly, the Max$_{\text{short}}^{\text{SS}}$ variant, in which the C-terminals of the leucine zipper are linked by a disulfide bond, unfolds with a specific unfolding enthalpy of 20.1 $\pm$ 0.2 J g^{-1} (at 43 °C; see Figure 3). This observation is in accord with the suggestion that the C-terminal part of the leucine zipper is partly or completely unfolded (*16*). The disulfide link tethers the frying (or unfolded) C-termini and assists folding (coiled coiling) along the entire length of the leucine zipper, an effect manifested with heat release.

Unfolding Energetics. Although essentially unstructured and not participating in packing interactions with either each other or with the b-HLH-LZ domain, the "tails" confer significant stabilization to Max p21 (Figure 3). The molecular basis of stabilization is unclear at present, but it was hypothesized that the stabilization originates from electrostatic effects (*16*). Indeed, it was suggested that the abundance of negative charges in the N-terminal segment of Max p21 might help to overcome the electrostatic destabilization originating from charge−charge repulsion between clusters of positive charges at the HLH interface around helix H1 (*12*). At the melting temperature of 46.3 °C (125 μM monomer concentration) the calorimetric estimates of the unfolding parameters are $\Delta H_{\text{cal}} = 370 \pm 20$ kJ mol^{-1}, $\Delta S_{\text{cal}} = 1.1 \pm 0.1$ kJ K^{-1} mol^{-1}, and $\Delta C_{\text{p}} = 3.6 \pm 0.4$ kJ K^{-1} mol^{-1}. Using the temperature of maximal heat absorption ($T_{\max}$), the molar heat capacity at that temperature ($C_{p,\max}$), the unfolding heat capacity increment (ΔC_p), and the (assumed) molecularity of unfolding process ($n = 2$), the effective van't Hoff enthalpy can be calculated as $\Delta H_{\text{vH}} = (\sqrt{2} + 1)T_{\max}\sqrt{R(C_{p,\max} - (\Delta C_p\sqrt{n})/(\sqrt{n+1}))} = 400$ kJ mol^{-1}. Thus, the ratio $\Delta H_{\text{vH}}/\Delta H_{\text{cal}}$ is 1.07, in reasonable agreement with a two-state unfolding between folded dimer and unfolded monomer. The highly cooperative character of unfolding is further supported by the coincidence the unfolding parameters derived by calorimetry (this study) and thermal melting followed by CD spectroscopy (*16*).

Stability Curve of Max p21. The stability of Max p21 at 5, 20, and 25 °C was assessed from isothermal urea-induced unfolding experiments by following the change in ellipticity at 222 nm. The data could be modeled with a two-state transition between native and unfolded protein (Figure 4). As shown in Figure 5, the agreement between ΔG_{U} values thus obtained and data obtained from DSC experiments is excellent. At 37 °C, the dissociation constant is $K_{\text{D}} = 3.6 \times 10^{-6}$ M and compares well with the only one previous estimate of 7.1×10^{-6} M (*16*). For the isolated b-HLH-LZ domain K_{D} at 37 °C was 100×10^{-6} M, obtained from CD thermal melting (*17*). Fluorescence anisotropy assay yielded a K_{D} ~0.7×10^{-6} M at 20 °C (*13*). For comparison, our estimate for the stability of Max p21 at 20 °C is 0.004×10^{-6} M. These results are in further support of the pronounced stabilization effect the disordered extensions exert on the core b-HLH-LZ domain.

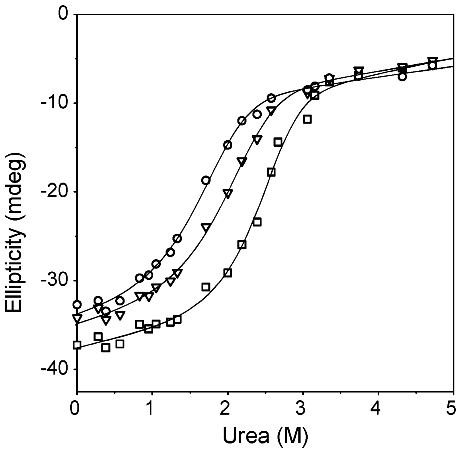


FIGURE 4: Urea-induced unfolding of Max p21. Experiments were performed with 30 μM protein (monomer equivalents) at 5 °C (squares), 20 °C (triangles), and 25 °C (circles). The solid lines are best fits according to a two-state unfolding model assuming equilibrium between folded dimer and unfolded monomer.

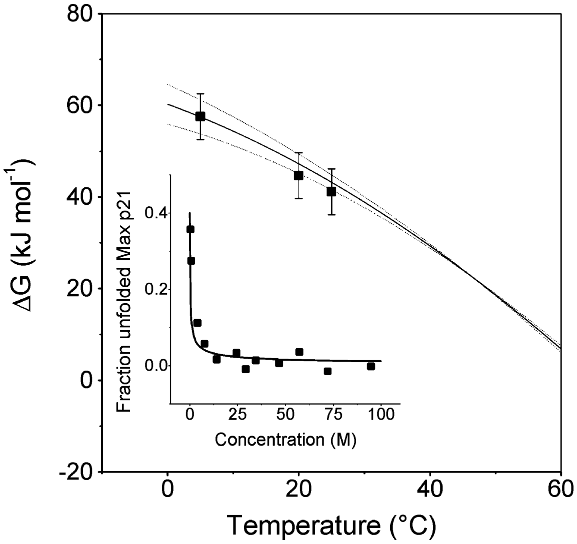


FIGURE 5: Thermodynamic stability of Max p21. The solid line represents the stability curve calculated from calorimetric data: $T_{\text{m}} = 46.3$ °C (125 μM monomer concentration), $\Delta H_{\text{cal}} = 370 \pm 20$ kJ mol^{-1}, and $\Delta C_p = 3.6 \pm 0.4$ kJ K^{-1} mol^{-1}. The dotted lines are the upper and lower error margins introduced by the experimental uncertainty. The symbols are ΔG measured by urea-induced denaturation at 5, 20, and 25 °C (data in Figure 4). Inset: Fraction unfolded Max p21 at 25 °C as a function of the protein concentration. The solid line is the function calculated as $f_{\text{U}} = (-K_{\text{D}} + \sqrt{K_{\text{D}}^2+8K_{\text{D}}M})/(4M)$, where $K_{\text{D}} = \exp[-\Delta G/(RT)]$ and M is the concentration in monomer equivalents. The symbols are the experimentally observed changes in ellipticity.

The stability curve (Figure 5) was used to simulate the fraction of unfolded protein as a function of the protein concentration (solid line, inset of Figure 5). The simulation was verified by measuring the change of MRE$_{222}$ (and the fraction unfolded) as a function of the total protein concentration (symbols in the inset of Figure 5). As one can notice, the experiment is in excellent agreement with the simulation, and demonstrates consistently that the dimeric state is populated to more than 97% for protein concentrations down to 10 μM between 5 and 25 °C. Hence, binding experiments were designed in this temperature and protein concentration range. Finally, the stability of Max p21 is insensitive to mutations in the basic region. Indeed, the midpoints of thermal denaturation measured by CD spectroscopy for all 7 variants studied here were within ±1 degree (not shown).

Thermodynamics of DNA Binding by Max p21

Experimental Design. Protein−DNA binding experiments are typically performed by titrating protein into DNA solution placed in the calorimetric cell. In cases of high affinity binding, low reagent concentrations are required to collect data allowing reliable calculation of the association constant and the binding enthalpy from the same binding isotherm (*22*). If the protein placed in the injection syringe is composed of noncovalently associated subunits, care should be taken that the protein concentration after dilution in the cell is always high enough, so that no dissociation takes place. Unfortunately, the thermodynamic stability of Max p21 is relatively low and a significant shift of the monomer−dimer equilibrium occurs at subsaturating concentrations (0.1−5 μM; see Figure 5). In contrast, the stability of the 21 bp DNA duplex used in this study is very high. From DSC data (duplex unfolding enthalpy was 570 kJ mol^{-1} at 80° for melting of 17 μM duplex), the duplex dissociation constant is on the order of 10^{-19} M and, therefore, no strand dissociation takes place down to low sub-micromolar concentrations. For this reason the ITC titrations were performed by titrating E-box duplex DNA into Max p21. To verify that the choice of titrant and titrand does not influence the results, control experiments of titrating protein into DNA or titrating DNA into protein were performed with the monomeric $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$ protein. The results were identical within error. However, this setup does not overcome the problem that there is a redistribution of the population of dimer and monomer during the titration experiment, since the cell contents become progressively diluted, and the free Max p21 concentration decreases as the saturation of DNA binding sites increases. As justified in the Supporting Information, due to the fortuitous combination of Max p21 dissociation constant, Max p21 dissociation enthalpy, and the large enthalpy change characterizing the formation of the Max p21/DNA complex, the heat effect of the monomer−dimer transition concomitant to binding is relatively small and influences the shape of the binding isotherm only slightly. The error from neglecting the monomer−dimer redistribution is not larger than the typical error caused by the uncertainty in concentration determination and the instrumental peak repeatability. The systematic overestimation of the molar binding enthalpy is ∼6 kJ mol^{-1}, as compared to a standard deviation of 6 kJ mol^{-1} calculated from triplicate experiments performed with different protein and DNA batches. The “distorted” binding constants are within the usual error range ±10−50%. We demonstrate below that the reported binding constants from ITC are in a very good correspondence with data derived from independent experiments. While, strictly speaking, the reported thermodynamic parameters are apparent, they are nevertheless reliable. It is important to note that the binding energetic differences reported next for the discussed mutants are not affected. Indeed, all Max p21 mutant proteins exhibit the same thermodynamic stability.

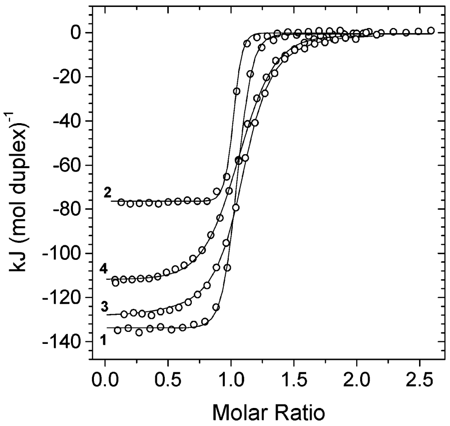


FIGURE 6: Examples of calorimetric isotherms describing binding of Max p21 to E-box DNA. The heats measured at each titration step were normalized for the molar concentration and corrected for unspecific heats (symbols). The continuous lines visualize best fits according to a 1:1 binding model. The figure illustrates the temperature variation of the binding enthalpy (compare trace 1 measured at 25 °C and trace 2 measured at 8 °C) and the decrease in affinity caused by addition of salt at 25 °C (total salt concentration was 150 mM (trace 1), 400 mM (trace 3), and 500 mM (trace 4)).

Thermodynamic Profile of the Wild Type Max p21/E-Box Complex. Examples of binding isotherms measured by ITC are shown in Figure 6. Addition of aliquots of E-box DNA duplex to Max p21 produces exothermic heat effects, which saturate as the molar ratio of duplex to Max p21 dimer increases. The stoichiometry in the data set was 0.98 ± 0.07 Max p21 dimer per DNA duplex (mean ± SD of more than 50 experiments, including experiments with Max p21 variants), in agreement with a 1:1 protein−DNA complex seen in the crystal structures of Max p21/E-box, c-Myc/Max/E-box, and Mad/Max/E-box complexes. There are suggestions that members of the Myc/Mad/Max family can form dimers of dimers via association patches provided by the leucine zipper domains (*23*). Indeed, tetrameric c-Myc/Max and Max/Max (but not Mad/Max) species binding two copies of the cognate DNA were observed by X-ray crystallography (*7*, *9*). However, ultracentrifugation experiments performed with concentrations typically used in ITC experiments at our experimental conditions failed to detect the presence of tetramers (see Figure S1 in the Supporting Information). Therefore, the measured thermodynamic parameters truly reflect the energy changes promoting formation of the Max p21 dimer to the E-box site.

Analysis of the binding isotherms obtained in standard buffer conditions and between 5 and 25 °C reveals that the association is enthalpy-driven and is opposed by entropy (Table 1). In this temperature range binding is too strong to obtain reliable estimates of the association constant, K_{A}, with reactant concentration in the low micromolar range. However, the binding enthalpy, ΔH_{A}, and its temperature dependence, $\Delta C_{p,\mathrm{A}}$, are defined with high precision. The same ΔH_{A} was measured in phosphate and Tris buffers

Table 1: Thermodynamic Parameters Characterizing Max p21 and Max p21 Variants Binding to E-box DNA Measured by ITC[a]

protein	$K_A \times 10^{-7}$ (M^{-1})	ΔG_A (kJ mol^{-1})	ΔH_A (kJ mol^{-1})	$T\Delta S_A$ (kJ mol^{-1})	ΔS_A (J K^{-1} mol^{-1})	$\Delta C_{p,A}$ (kJ K^{-1} mol^{-1})
Max p21	38.00 69.00[b]	−48.9	−137	−88	−295	−3.8
Max p21VL	610.00	−55.8	−140	−84	−282	nd[c]
Max_{short}^{SS}	480.00	−55.2	−129	−74	−248	−3.8
Max $p21^{H18A}$	2.30	−42.0	−124	−82	−275	−3.8
Max $p21^{N19A}$	180.00	−52.8	−133	−80	−268	−4.2
Max $p21^{E22A}$	2.15	−41.8	−126	−84	−282	−3.6
Max $p21^{R23A}$	2.23	−41.9	−130	−88	−295	−3.6
Max $p21^{R25A}$	0.22	−36.2	−125	−89	−300	−3.3
Max $p21^{R26A}$	$<10^{-3}$	>-25.0	nd	nd	nd	nd
Max $p21^{R50A}$	0.55	−38.5	−159	−119	400	−4.1

[a] All experiments were performed at 25 °C in 100 mM Na-phosphate, 100 mM KCl, pH 6.8, unless otherwise indicated. The standard error of K_A, δK_A was maximum 60%. Thus, the maximum error of ΔG_A is $\delta\Delta G_A = RT(\delta K_A/K_A) = 1.5$ kJ mol^{-1}. The maximum error of ΔH_A, $\delta\Delta H_A$, was on the order of 6 kJ mol^{-1}. The error of $T\Delta S_A$ is $\delta T\Delta S_A = \sqrt{(\delta\Delta H_A)^2+(\delta\Delta G_A)^2} = 6.5$ kJ mol^{-1}. The error of $\Delta C_{p,A}$ was estimated as ±0.2 kJ K^{-1} mol^{-1} by jackknife tests. [b] From DSC data according to eqs 1−3. [c] Not determined.

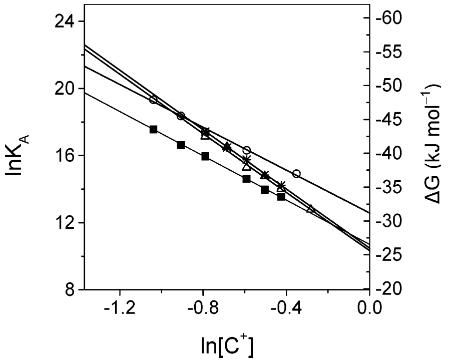


FIGURE 7: Influence of salt concentration on the affinity of Max p21 variants to E-box DNA. The binding constants were measured by ITC at 25 °C in increasing concentrations of KCl. Filled squares, wild type Max p21. Circles, Max $p21^{N19A}$. Triangles, Max_{short}^{SS}. Asterisks, Max p21VL. The lines are linear fits according to eq 4. Extrapolation to the left *y*-axis yields $K_A = e^{\ln K_A}$ at the standard buffer conditions (0.154 mM cation concentration), which are listed in Table 1. Extrapolation to the right *y*-axis yields the nonelectrostatic contribution to ΔG_A (ΔG^{ne}) estimated at 1 M cation concentration. The slopes of the straight lines equal the product $Z\psi$ as defined in eq 4.

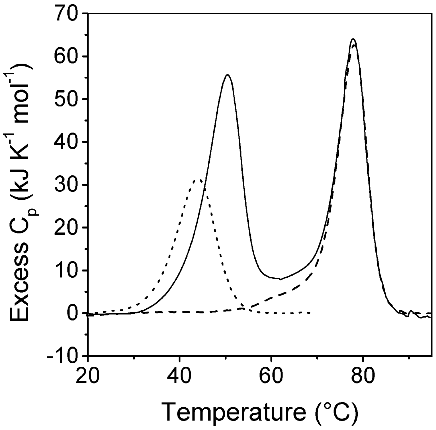


FIGURE 8: DSC melting profiles of Max p21 (dotted line), E-box DNA (dashed line), and the 1:1 protein−DNA complex (continuous line). The concentration of all species was 17 μM (protein dimer, DNA duplex, and complex).

having different heats of protonation. Hence, there is no (net) proton release/uptake accompanying binding.

In order to get an accurate estimate of K_A, we explored the well-known sensitivity of protein−DNA complexes to the concentration of cations (*24, 25*). Binding experiments were performed in increasing concentrations of KCl at 25 °C. As illustrated in Figure 6 the binding isotherms become increasingly shallower at higher salt concentration because binding gets weaker (by a factor 55 from 350 mM to 650 mM cation concentration). Reliable K_A can be calculated from such isotherms. As in many documented cases, in the absence of specific anion effects, ln K_A depends linearly on the logarithm of cation concentration (ln [C^+]) in the range of salt concentrations where the water activity is negligibly perturbed by the presence of salt (Figure 7). Linear extrapolation to our standard buffer conditions yields K_A(25 °C) = (3.8 ± 0.2) × 10^8 M^{-1} (K_D = (2.7 ± 0.1) × 10^{-9} M).

Whereas changes of K_A with temperature, pH, and pressure have a rigorous thermodynamic basis, this is not the case for the effect of salt. To verify the reliability of the linear extrapolation, DSC experiments were designed to obtain an independent estimate of K_A. To this end, DSC melting traces were recorded with Max p21, E-box DNA duplex, and the 1:1 Max p21/E-box complex. The midpoints of thermal unfolding of Max p21 and E-box DNA are separated by ~35 °C (Figure 8). Melting of the 1:1 complex produces two peaks. The first one overlaps partially with the thermogram of isolated Max p21 and corresponds to cooperative protein−DNA complex melting, which is intimately coupled to Max p21 unfolding. The ~7 °C shift in the temperature of maximum heat absorption relative to the maximum of the Max p21 peak reflects the stabilization of the protein caused by protein−DNA contacts. Melting of the complex releases free DNA duplex, which melts at a much higher temperature (the second peak of the continuous trace in Figure 8). Using the known unfolding enthalpy (ΔH_P) and unfolding heat capacity change ($\Delta C_{p,P}$) of the Max p21 at the melting temperature (T_P) in the absence of DNA, the association constant K_A at the temperature corresponding to half-completion of complex melting (T_C) can be calculated according to the following equation (*26*):

$$K_A(T_C) = \frac{\exp\left\{-\frac{\Delta H_P}{R}\left(\frac{1}{T_C}-\frac{1}{T_P}\right)+\frac{\Delta C_{p,P}}{R}\left(\ln\frac{T_C}{T_P}+\frac{T_P}{T_C}-1\right)\right\}-1}{[D]_{T_C}} \quad (1)$$

The term $[D]_{T_C}$ in the denominator of eq 1 is the concentration of free DNA duplex at T_C. Since binding is strong, the population of the 1:1 complex is higher than 0.98 at the onset of heat absorption (~30 °C) and, therefore, $[D]_{T_C} = [D]_{tot}/2$. The calculated K_A(50 °C) is $(1.9 \pm 0.3) \times 10^6$ M^{-1}. This number compares well with $K_A(50\ °C) = (1.3 \pm 0.5) \times 10^6$ M^{-1} obtained by using the parameters directly measured by ITC at 25 °C according to

$$K_A(T_C) = K_A(25\ °C)\exp\left\{-\frac{\Delta H_A(25\ °C)}{R}\left(\frac{1}{T_C}-\frac{1}{298.15}\right)+\frac{\Delta C_{p,A}}{R}\left(\ln\frac{T_C}{298.15}+\frac{298.15}{T_C}-1\right)\right\} \quad (2)$$

The data from ITC (ΔH_A and $\Delta C_{p,A}$) and DSC (K_A(50 °C)) can be combined to yield an estimate of K_A at 25 °C:

$$K_A(25\ °C) = K_A(50\ °C)\exp\left\{-\frac{\Delta H_A}{R}\left(\frac{1}{298.15}-\frac{1}{323.6}\right)+\frac{\Delta C_{p,A}}{R}\left(\ln\frac{298.15}{323.6}-\frac{298.15}{323.6}+1\right)\right\} \quad (3)$$

The resulting K_A is $(6.9 \pm 2) \times 10^8$ M^{-1}, in close agreement with the value obtained by linear extrapolation of the data collected at high-salt conditions ($(3.8 \pm 0.2) \times 10^8$ M^{-1}). The consistency of the binding constants in an extended temperature interval allows a reliable estimate of the stability of the Max p21/E-box complex at 37 °C: $(4 \pm 1) \times 10^7$ M^{-1} ($K_D = (25 \pm 6) \times 10^{-9}$ M). To our knowledge, this is the first direct estimate of the binding affinity of the complete gene product of Max p21 to E-box at the physiologically relevant temperature. Previously, the affinity of the b-HLH-LZ to E-box containing DNA duplexes was estimated to be in the range 1 and 30 nM at temperatures between 20 and 37 °C and at similar buffer conditions (*13*, *14*). It follows that while the protein N-terminal and C-terminal regions outside the b-HLH-LZ contribute to the stability of the p21 dimer, they are not involved in any significant interactions with DNA.

The thermodynamic parameters describing formation of the Max p21/E-box complex at 25 °C are $\Delta G_A = -48.9 \pm 2.0$ kJ mol^{-1}, $\Delta H_A = -137 \pm 8$ kJ mol^{-1}, $\Delta S_A = -0.295 \pm 0.025$ kJ K^{-1} mol^{-1} ($T\Delta S_A = -88 \pm 8$ kJ mol^{-1}), and $\Delta C_{p,A} = -3.8 \pm 0.2$ kJ K^{-1} mol^{-1} (Table 1). Both ΔH_A and $T\Delta S_A$ are large numbers, which cannot be interpreted as the enthalpy accumulated in intermolecular bonds, or as the entropy change associated with these bonds. Rather, the total binding parameters contain the energetic expenditures for the conformational transition of the basic region from the largely unfolded, free state to its α-helical, bound state. Considering ΔH_A, the "genuine" binding enthalpy can be estimated by subtracting the enthalpy of α-helix formation from the total observed enthalpy change. Fourteen residues per Max chain undergo coil-to-helix transition (*12*). The enthalpy of helix formation was estimated as −2.5 to −3.8 kJ mol $residue^{-1}$ depending on the chemical identity of the side chain involved (*27*), and the weight of evidence suggests −4 kJ mol $residue^{-1}$ as the upper limit for a typical hydrogen bond (*28*). Combining these numbers, the "genuine" enthalpy stabilizing the Max p21/E-box complex can be estimated as −25 to −40 kJ mol^{-1}. Assuming that formation of 10 hydrogen bonds, as seen in the crystal structure, represents the major enthalpic contribution to binding, the mean enthalpic content of such bonds is −2.5 to −4.0 kJ mol^{-1}, in agreement with the estimated enthalpic content of hydrogen bonding cited just above. A very similar picture was recently suggested for DNA recognition by the b-ZIP domain of GCN4, on the basis of direct measurement of the enthalpic contribution for folding of the similarly long basic region (*29*). The large unfavorable entropy change indicates that the loss of conformational entropy dominates the entropic benefits from dehydration of molecular surface and the polyelectrolyte effect. According to a back-on-the-envelope calculation the sum of the entropic contributions (in J K^{-1} mol^{-1}) arising from (i) immobilization of 22 side chains becoming buried by ≥20 $Å^2$ in the complex (−300; ref (*30*)), (ii) immobilization of the backbone of 28 residues (−420; ref (*31*)), (iii) dehydration of molecular surface (+310, considering only the protein−DNA interface dehydration; (*32*)), (iv) polyelectrolyte effect (+120; ref (*25*)), and (v) loss of translational and rotational degree of freedom (−35: ref (*33*)) predicts a total predicted entropy change of −325 J K^{-1} mol^{-1}, as compared to the experimental value of −295 J K^{-1} mol^{-1}. The calculation is indeed very crude, yet it might give some idea about the balance of entropic factors involved in complex formation. From the reduction of solvent accessible surface, one can estimate that the dehydration of the protein−DNA interface will cause a heat capacity decrease of approximately 0.4 J K^{-1} mol^{-1}, ten times lower than the experimental value (−3.8 J K^{-1} mol^{-1}). However, Dragan et al. convincingly demonstrated that the similar discrepancy observed for GCN4 b-ZIP binding to DNA is caused by the temperature-induced changes in the structural content (and the extent of DNA-induced refolding for that matter) of the basic region in the temperature range of ITC experiments (*29*). In principle, the effect can be quantified by comparing the temperature dependence of the heat capacities measured for the associated and dissociated state of the protein−DNA complex (*34*, *35*). Unfortunately, the limited solubility of Max p21 precluded precise heat capacity measurements in sufficiently concentrated solutions. Nevertheless, the sum of the heat capacities of the isolated components is significantly higher than the heat capacity of the Max p21/DNA complex and the difference strongly depends on the temperature. As explained in detail in the Supporting Information, from the available data we estimate that the temperature-induced shift of the coil-to-helix equilibrium of the basic region and any other temperature-dependent structural changes collectively contribute ~−3 kJ mol^{-1} K^{-1} to the apparent ΔC_p. Hence, ΔC_p arising from intermolecular interactions and dehydration of the surface is ~−1 kJ mol^{-1} K^{-1}, or even smaller (see above).

Number of Cations Released upon Binding and Electrostatic Contribution to ΔG_A. The linear dependence between ln K_A and ln $[C^+]$ allows estimation of the number of cations released from the phosphate backbone upon protein binding, i.e., the number of protein−phosphate backbone ionic contacts (*24*, *25*). Furthermore, the electrostatic and non-electrostatic components of the total binding free energy can be estimated (*29*). For the equilibrium process

$$\mathrm{P} + \mathrm{D} + n\mathrm{C}^{+} \rightleftarrows \mathrm{PD} + m\mathrm{C}^{+}$$

the following equation holds (*25*):

$$\ln K_{\mathrm{A}} = \ln K_{\mathrm{A}}^{ne} - Z\psi \ln\,[\mathrm{C}^{+}] \qquad (4)$$

P, D, and PD represent the equilibrium concentrations of protein, DNA, and protein−DNA complex, respectively; C^{+} is the molar concentration of cations; n and m are the stoichiometric coefficients. $Z = (n - m)$ represents the number of protein−DNA ionic contacts, and $\psi = 0.64$ is the number of released cations per phosphate group (*36*). From the slope of the plot according to eq 4, $Z = 10.3$. Therefore, the estimated number of ionic contacts bridging Max p21 to the E-box phosphate backbone is 5 per Max p21 chain. Crystallographic analysis identified the existence of 6 such contacts (*7*). Given the quality of the experimental data, the discrepancy is unlikely caused by experimental uncertainties. Rather, we presume that one of the suggested contacts is not realized in solution at the selected conditions. Incidentally, the same sensitivity on the cation concentration as the wild type complex was observed for the Max $\mathrm{p21}^{\mathrm{N19A}}$ mutant ($Z = 9.9$; see Figure 7). Since Asn 19 is one of the side chains implicated in electrostatic interaction with the phosphate backbone, it appears plausible to assume that the contact seen in the X-ray structure is not populated in solution. Interestingly, the closest distance of Asn 19 (via OD1) to the DNA backbone is larger than 3.8 Å in the structure of the Max b-HLH-LZ/DNA complex (1AN2; (*8*)). In fact, the replacement of the Asn19 side chain by alanine stabilizes the complex (see below), which is difficult to comprehend if favorable polar interactions with DNA occur at that site. The conclusion is supported by analysis of the total binding free energy in terms of electrostatic and nonelectrostatic components. The nonelectrostatic component of ΔG_{A} can be roughly estimated from $-RT \ln K_{\mathrm{A}}$ according to eq 4 in the limit of $[\mathrm{C}^{+}] = 1$ M, since the second term of eq 1 is zero. Although it is not possible to estimate to what extent electrostatic effects are attenuated at this salt concentration, it appears that they contribute almost half of the binding free energy measured for Max p21 binding at 25 °C ($\Delta G^{\mathrm{el}} \sim -22$ kJ mol^{-1}; see Table 1 and Figure 7). The electrostatic contribution to ΔG_{A} of the Max $\mathrm{p12}^{\mathrm{N19A}}$/DNA complex is identical within error with that of the wild type complex, indicating that the affinity increase caused by the N19A mutation is governed by nonelectrostatic effects (ΔG^{nel}, Table 2).

Stabilization of the Leucine Zipper Domain Increases the DNA-Binding Affinity of Max p21. The Origin of the Phenomenon Is Electrostatic. Spectroscopic data have shown that the midpoint of thermal dissociation of the Max b-HLH-LZ/DNA complex is shifted to higher temperature if LZ of Max is stabilized (*16*). However, the origin of this effect has never been pursued in detail. Now we provide hints about the molecular origin of the phenomenon. As seen in Table 1, either introduction of an S−S bridge at the C-terminus of the leucine zipper of the core b-HLH-LZ domain, or the double N68V/H71L mutation in the context of the full length Max p21 increases K_{A} by a factor of 12−15. Both variants are significantly more stable than the wild type counterparts (*16*, *17*) (see also Figure 3 to compare the stability of Max p21 and $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$). From the data taken at face value from Table 1 it would appear that stabilization of the leucine zipper favors binding entropically, especially when the chains of b-HLH-LZ are covalently linked in the $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$ variant. Still, the observed decrease in ΔH_{A} off-setting the entropic benefit is not easy to understand. In Max p21VL better binding is accomplished by simultaneous action of enthalpic and entropic factors, yet the enthalpy−entropy balance remains elusive, since the differences are clearly within the experimental error margins. There are reasons to hypothesize that DNA contacts from the loop region and/or helix H1 are created or become stronger in the stabilized versions of Max b-HLH-LZ. Indeed, the more stable complexes formed by Max p21VL and $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$ are also more sensitive to the concentration of cations. Formation of both complexes is accompanied by the release of 7 cations per binding site ($Z = 13.8$ for Max p21VL and $Z = 13.7$ for $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$; see Table 2 and Figure 7). It follows that some additional electrostatic contacts in these variants increase the DNA-binding affinity and these new contacts are restricted to side chains within the b-HLH-LZ domain. It turns out that ΔG^{el} is significantly larger in the case of Max p21VL and $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$ binding ($\Delta G^{\mathrm{el}} \sim -30$ kJ mol^{-1}). In fact, the observed increase in ΔG^{el} completely explains the stronger binding of Max p21VL and $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$, since the nonelectrostatic component of ΔG_{A}, $\Delta G^{\mathrm{nel}} = \Delta G_{\mathrm{A}} - \Delta G^{\mathrm{el}}$, is identical within error for Max p21, Max p21VL, and $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$.

Table 2: Number of Cations Released upon Formation of the Max p21/DNA Complex and Estimation of the Electrostatic and Nonelectrostatic Contribution to ΔG^{a}

protein	Z^{b}	$\Delta G^{\mathrm{nel}\ c}$	$\Delta G^{\mathrm{el}\ d}$	$\Delta G^{\mathrm{nel}\ d}$
Max p21	10.3 ± 0.1	-48.9 ± 0.2	-22.4 ± 0.3	-26.5 ± 0.2
Max p21VL	13.8 ± 0.1	-55.8 ± 0.2	-29.9 ± 0.3	-25.9 ± 0.2
$\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$	13.7 ± 0.1	-55.2 ± 0.2	-29.7 ± 0.3	-25.5 ± 0.2
Max $\mathrm{p21}^{\mathrm{N19A}}$	9.9 ± 0.1	-52.8 ± 0.2	-21.6 ± 0.3	-31.2 ± 0.2

[a] At 25 °C, pH 6.8, ΔG values in kJ mol^{-1}. [b] Calculated as $Z = -\mathrm{d} \ln K_{\mathrm{A}}/\mathrm{d}[C^{+}]/0.64$. [c] Nonelectrostatic contribution to ΔG_{A} estimated from $-RT \ln K_{\mathrm{A}}$ at 1 M cation concentration. [d] Electrostatic contribution to ΔG_{A} estimated from $\Delta G^{\mathrm{el}} = \Delta G_{\mathrm{A}} - \Delta G^{\mathrm{nel}}$.

Altogether, the data strongly suggest that the Asn 19 side chain is not involved in electrostatic contacts with the DNA backbone, while additional protein−phosphate backbone interactions are responsible for the higher-affinity binding of the variants with stabilized b-HLH-LZ domain. We are reluctant to embark on lengthy speculations as to which side chains may be involved. In principle, there are many highly conserved basic residues in close proximity to the DNA backbone, involving the presumably quite unrestrained stretch Lys 13-Arg 14-Arg 15 preceding the “canonical” portion of the basic region (His 18 to Arg 26), Lys 24 in the middle of the basic region, Lys 30 from helix H1, as well as Lys 47 from the loop region. Indeed, some of these interactions were described in or can be surmised from the crystal structures of the c-Myc-Max−E-box and $(\mathrm{Max})_2$−E-box complexes, both containing only the b-HLH-LZ protein domains (*8*, *9*). The feasibility of these contacts is less obvious in the only available structure of the Max p21/E-box complex. It is perhaps most natural to assume that a more stable, and possibly less fluctuating leucine zipper alters the structural or dynamic properties, or both, of the adjacent HLH domain. It has been argued that stabilization of structural scaffolds or stabilization of local conformations could improve the DNA-binding affinity of transcription factors (*37*, *38*). Possibly, conformational rearrangement of the loop region and slight positional reorientation of helices H1 and H2 facilitate new protein−DNA contacts. In fact, the NMR structure of the free, disulfide linked Max b-HLH-LZ domain bearing the N68V/H71L double mutation indicates a loop conformation facilitating direct contact of Lys 47 and Ser 49 to DNA phosphates (*12*). Lys 47 points away from the duplex in $(\mathrm{Max})_2$/E-box X-ray structure. As another, very likely candidate we suggest Lys 30 of helix H1, which forms very well-defined charge−charge contacts with DNA phosphates (distance <3.5 Å) in both monomers of the c-Myc-Max/E-box complex.

Mutational Analysis of the Max p21/E-Box Interface

On the basis of structural information six residues from the highly conserved basic region of the Max p21 monomer contact the Cyt(1)-Ade(2)-Cyt(3)-Gua(4)-Thy(5)-Gua(6) E-box core sequence (*7*). (Although more contacts could be inferred from the X-ray structures of $(\mathrm{Max})_2$/E-box, c-Myc-Max/E-box and Mad-Max/E-box complexes formed by the corresponding b-HLH-LZ core domains, we analyze the contacts as they are defined in the only available complex of the full length Max p21.) Residues His 18, Glu 22, and Arg 26 are involved in specific hydrogen bonds to DNA bases. Mutation of Glu 22 abolishes discrimination between E-box and nonspecific DNA (*39*). The contact formed by Arg 26 to the central guanine crucially governs the specificity of b-HLH-LZ proteins for class B E-box elements (*40*). Arg 26 is anchored also to the DNA backbone since both NH2 and Nϵ atoms are positioned very close to phosphate oxygen atoms. Nonspecific interactions with the phosphate backbone are formed by Asn 19, Arg 23, and Arg 25. Arg 50, which is located at the start of helix H2, makes a main-chain and a side-chain contact with backbone phosphates. A schematic representation of the interactions is shown in Figure 1. To probe for the energetic role of particular contacts we replaced the corresponding side chains by alanine and determined the energetic signature of the mutant complexes by ITC. The measured thermodynamic parameters at 25 °C are listed in Table 1.

Figure 9 illustrates the changes in the energetic profile of the mutant Max p21/E-box complexes. The bars represent $\Delta\Delta G$, $\Delta\Delta H$, and $T\Delta\Delta S$ values ($\Delta\Delta X = \Delta X^{\mathrm{mutant}} - \Delta X^{\mathrm{wild\ type}}$) calculated for binding of the Max p21 dimer to E-box duplex. We would like to remind the reader that the indicated energetic differences are the combined effect of two simultaneous mutations, one on each monomer. We assume that the structural symmetry seen in the crystal structure translates into energetic equality of the contacts realized in the two half sites. Removal of the base-specific hydrogen bond of Arg 26 by replacing the central E-box guanine by cytosine in one or both E-box half sites has demonstrated essentially no cooperativity (*6*). This result should not be taken as ultimately proving the independence of the half sites since all other protein−DNA contacts have been left intact in the cited study. Small distance differences between interacting groups might result in serious energetic differences. However, we are not aware of any study documenting cooperativity between half sites in DNA recognition by (structurally) symmetrical homodimer proteins.

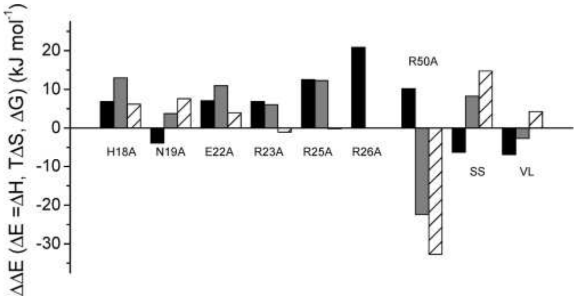


FIGURE 9: Changes in the energetic profile of the Max p21/E-box complex upon alanine substitution or LZ stabilization. The bars represent the changes in ΔG_{A} (black), ΔH_{A} (gray), and $T\Delta S_{\mathrm{A}}$ (hatched) calculated as $\Delta\Delta E = \Delta E^{\mathrm{mutant}} - \Delta E^{\mathrm{wildtype}}$ (numerical ΔE values from Table 1). $\Delta\Delta G > 0$ indicates that the mutation is destabilizing. Mutations are destabilizing the complex enthalpically if $\Delta\Delta H > 0$. Entropically favorable mutations are manifested by $T\Delta\Delta S > 0$. All experiments were performed at 25 °C in the standard buffer.

Three mutations, H18A, E22A, and R23A, exhibit modest decreases of affinity (∼sixteen time increase in K_{D}). The effect of removal of Arg 25 and Arg 50 is more pronounced (a K_{D} increase of 160 and 60 times, respectively). Two mutations stand out. The wild type Asn 19 side chain appears to destabilize the complex relative to alanine. Removal of the Arg 26 side chain beyond the C_{β} atom is extremely destabilizing. We cannot give a very precise estimate for the decreased affinity of the Max p21$^{\mathrm{R26A}}$ protein. ITC experiments were not possible at concentrations sufficiently high as to allow precise measurement of the binding constant at 25 °C. The observed heat effect of binding was small as well. These experiments indicated ∼50 μM as the lower bound of K_{D}. Since the pronounced reduction of heat release relative to wild type binding introduced ambiguity in the ITC results, another estimate of the Max p21$^{\mathrm{R26A}}$ affinity was obtained by CD spectroscopy. Recently, it was convincingly demonstrated that the decrease in MRE_{222} upon titration of Max p21 with increasing amounts of DNA directly reflects the population of the protein/DNA complex (*41*). Mixing of equimolar (50 μM) Max p21$^{\mathrm{R26A}}$ and E-box duplex results in a decrease in MRE_{222} of 400 deg cm dmol^{-1}. This is ∼15% of the average MRE_{222} decrease observed at full saturation (3000 deg cm dmol^{-1}). With this information, K_{D} ∼200 μM was calculated. Irrespective of the exact K_{D} value, it is clear that the R26A mutation destabilizes the complex by a factor of at least 10^4 (K_{D}).

Is it possible to rationalize the energetic effect of the mutations? Visual inspection of the protein/DNA cocrystal structure provides clues for some sites. Removal of the four base-specific and buried hydrogen bonds formed by Glu 22 to N4 of Cyt(1) and N6 of Ade(2) (two hydrogen bonds per half site) is energetically not very costly: each bond contributes 1.7 kJ mol^{-1} on average to ΔG_{A}. Interestingly, if the bonds are broken by mutating out the hydrogen donors, the average contribution per hydrogen bond is more than twice larger, 3.7 kJ mol^{-1} (*41*). It is likely that the carboxyl group of Glu 22 experiences large unfavorable dehydration, which is uncompensated in the absence of hydrogen bond partners. The problem appears relieved in the Max $\mathrm{p21}^{\mathrm{E22A}}$ mutant lacking the side chain carboxylate. The dramatic destabilization caused by the Arg 26 replacement is surprising at first glance, since the contribution of each base-specific hydrogen bond between NH1 of Arg 26 and N7 of Gua(4) has been recently estimated as 2−3 kJ mol^{-1} (*6*). It should be noted, however, that the Arg 26 side chain is engaged additionally in two sugar−phosphate backbone contacts, which are also lost (broken) in the R26A mutant studied here. Apart from the energetic contribution of each sugar−phosphate backbone contact *per se* (which can be substantial as seen in Figure 9) such interactions might help to fix the arginine side chain in a position for optimal hydrogen bonding and certainly help to overcome the entropic loss from freezing of the side chain. In any case, Arg 26 represents a crucial “hot spot” at the binding interface.

Overall, there is a large variation in the destabilizing effect of the mutations. On average, backbone contacts contribute more to stabilization of the complex than base-specific hydrogen bonds. This is in line with previous studies with X-to-Ala mutations showing that substitution of residues making contacts with the sugar−phosphate backbone may produce mutant proteins more severely defective in DNA binding than substitution of residues making base-specific contacts (e.g., (*42*)). The energetic content of such “unspecific” interactions might also differ considerably. A clear example is presented by Arg 23 and Arg 25, the effect of the latter being twice larger. In fact, Arg 23 is oriented much more favorably toward the phosphate oxygen atoms than Arg 25. However, it is also significantly more exposed to the solvent than Arg 25. Moreover, we note that the Arg 25 Nϵ atom faces the carboxylic group of Glu 22 (it is <3.5 Å away from both carboxylate oxygens). This observation suggests that the Arg 25 side chain may help in optimizing the geometry of the base-specific contacts of Glu 22. If so, it exerts an “indirect” additional stabilization of the complex, which is lost upon mutation to alanine. Particularly intriguing is the role of Asn 19. The residue is highly conserved within the Myc/Max/Mad network of transcriptional factors and it closely approaches the DNA backbone according to structural data, yet the removal of this contact by mutating out the asparagine amide group increases the affinity for DNA. According to the original annotation of the available coordinate files the contact to the phosphate oxygen is made by either OD1 or ND2 atoms of Asn 19 (distance <3 Å). Since the assignment of the electron density to OD1 or ND2 atoms in not very highly resolved structures is ambiguous, we have attempted a simple visual evaluation of the possibility that the potentially unfavorable Asn OD1-to-O1P contact can be relieved by rotating the C_β−C_γ bond to position ND2 close to O1P. No clear conclusions could be drawn. It appears that the environment of the Asn 19 amide group is electrostatically intensive due to the close proximity of groups bearing partial positive charge (ND1 of His 18) or partial negative charge (the carbonyl oxygen of Arg 15) and bound water molecule(s). We are currently investigating the problem by means of MD simulations.

Concerning the energetic partitioning of the changes in affinity, we cannot provide a detailed structural explanation of the observed enthalpy−entropy balance of particular mutations. First, the experimental error is sizable in comparison to the magnitude of the measured $\Delta\Delta H_{\mathrm{A}}$ and $\Delta\Delta S_{\mathrm{A}}$. Second and more important, binding is tightly coupled to (partial) coil-to-helix transition of the Max p21 basic region and to bending of the E-box duplex. In principle, the energetic signature of these processes could be affected by mutation. Third, replacement of larger side chains having polar groups by small and nonpolar alanine could have caused redistribution of water molecules at and near the binding interface. With these considerations in mind, in the following we briefly discuss some general trends and possible sources of the energetic signature of some mutations.

We first consider mutations in the basic region, excluding the R26A mutation, for which $\Delta\Delta H_{\mathrm{A}}$ and $\Delta\Delta S_{\mathrm{A}}$ could not be measured. The R50A replacement, which is outside the boundaries of the basic region, exhibits a completely different energetic signature and should be discussed separately. It is clear from Figure 9 that all mutations in the basic region are linked to loss of enthalpic interactions, including the stabilizing N19A replacement. For side chains stabilizing the protein−DNA complex, the magnitude of enthalpic destabilization upon mutation dominates over the entropic effect, but there is no correlation between the total destabilization ($\Delta\Delta G$) and its enthalpic ($\Delta\Delta H$) and entropic ($\Delta\Delta S$) components, as noted in diverse other macromolecular systems. Removal of a backbone contact could be enthalpically as costly as the removal of two hydrogen bonds (R25A *versus* H18A and E22A). The enthalpic contribution from formally identical backbone contacts could also be very different (R23A *versus* R25A). In fact, the enthalpy term completely determines the magnitude of stabilization provided by the latter two arginine side chains.

The changes in heat capacity upon mutation are quite small (5 to 13% of $\Delta C_{p,\mathrm{A}}$), within the experimental uncertainty (Table 1). It should be considered, however, that the total heat capacity decrement is largely dominated by the refolding of the basic region. Since single alanine mutation are not expected to alter the coil-to-helix transition of the basic region significantly, and the “rigid-body” heat capacity change is rather small (−1 kJ K^{-1} mol^{-1} or smaller), the measured $\Delta\Delta C_{p,\mathrm{A}}$ values become sizable. There is no correlation between the experimental $\Delta\Delta C_{p,\mathrm{A}}$ and the amount of polar, nonpolar, and total buried surface change caused by mutation. Interestingly, $\Delta\Delta C_{p,\mathrm{A}}$ strongly correlates with the free energy effect of the mutations ($\Delta\Delta G_{\mathrm{A}}$). The stabilizing N19A replacement is linked to the largest negative heat capacity increment (−0.4 kJ mol^{-1} K^{-1}), while the most destabilizing mutation (R25A) is accompanied by the largest positive $\Delta\Delta C_{p,\mathrm{A}}$ (0.5 kJ mol^{-1} K^{-1}). Overall, $\Delta\Delta C_{p,\mathrm{A}}$ and $\Delta\Delta G_{\mathrm{A}}$ are correlated by $R^2 = 0.96$. This observation possibly indicates a complicated balance of forces and mechanisms leading to change in affinity, since neither differential parameter (including $\Delta\Delta H_{\mathrm{A}}$ and $\Delta\Delta S_{\mathrm{A}}$) correlates with the change of surface burial upon mutation, and there are no other cross-correlations between $\Delta\Delta X$ parameters.

For all sites entropic factors disfavor the wild type side chains, or else are negligible. The entropy gain cannot be explained with the gain of side chain entropy when longer side chains are replaced by alanine, since the wild type arginine side chains appear not to be penalized entropically. We note the general trend that the favorable entropic effect

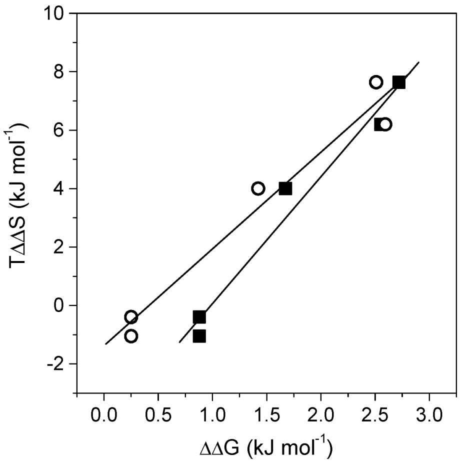


FIGURE 10: Correlation between the measured change in entropy upon X-to-Ala substitutions of side chains in the basic region of Max p21 and the α-helical propensity of the replaced side chains. $T\Delta\Delta S$ is plotted against $\Delta\Delta G$, which ranks the amino acid side chains according to their preference to adopt α-helical conformation (low $\Delta\Delta G$ indicates high α-helix propensity, relative to alanine, for which $\Delta\Delta G = 0$). The helix propensity scales derived entirely from experimental data with peptides and proteins are shown. Filled symbols, scale by Pace and Scholtz ($R^2 = 0.97$; ref (*43*)). Open symbols, scale AGADIR ($R^2 = 0.96$; ref (*44*)).

decreases the closer the site of mutation is to the point where the basic region emerges from helix H1. Sauvé et al. have demonstrated by NMR the existence of a persistent helical turn spanning residues Arg 25 to Ile 29 in the absence of DNA, even at 35 °C (*12*). Furthermore, the Arg 25 and Arg 26 side chains adopt an extended conformation, which is virtually identical to the conformation observed in the Max p21/E-box complex. It is therefore natural to assume that substituting either residue by alanine (having higher, yet very similar helical propensity as arginine) would not interfere with both the fractional population of the helical turn and the conformation of the adjacent arginine side chain. Differently, His 18, Asn 19, Glu 22, and Arg 23 obtain a stable α-helical conformation only in the context of the protein−DNA complex, but may exist in transient α-helical states also in the free protein. As far as stabilization of such states spanning the basic region in the free protein would decrease the entropic penalty for finding the proper conformation facilitating formation of intermolecular contacts, alanine mutation in the corresponding sites is expected to stabilize the complex entropically. Figure 10 illustrates the very strong correlation between the increase in α-helical propensity upon X-to-Ala mutation and the entropic benefit the mutation exhibits. Although alanine substitutions might diminish the entropic penalty, the *total* energetic effect of mutation is greatly modulated by the enthalpic factors (i.e., formation/breakage of bonds and the concomitant hydration changes). This point is nicely illustrated by comparing the energetic signature of H18A and N19A mutations.

The energetic effect of the R50A mutation is balanced in a radically different way. First, both the enthalpy and the entropy changes are much larger than the changes observed for all other sites. Second, the replacement is favored by a large enthalpic effect and is disfavored by an even larger entropic effect. Since the mutation eliminates only one of the two contacts of Arg 50 to the phosphate backbone (the second one being formed by the peptide amide), this means that the interactions formed by the Arg 50 side chain destabilize the wild type complex enthalpically, yet strongly promote DNA binding entropically. Both the magnitude and the sign of the observed energetic changes are very unlikely caused by elimination of a single contact to the phosphate backbone. Rather, we envisage contributions from structural rearrangements of the protein in response to the mutation. The methylene groups of the Arg 50 side chain pack against Ile 29 and Phe 33 and complement the HLH hydrophobic core. These interactions are present both in the X-ray structure of the Max p21-E-box complex and in the NMR ensemble of the free Max protein but are eliminated in the Max p21^{R50A} mutant. Two nonexclusive scenarios are possible. (i) Removal of the hydrophobic moiety of the Arg 50 side chain by alanine mutation causes loosening of the packing of the HLH hydrophobic core and promotes local unfolding. In the Max p21^{R50A}/DNA complex, the contact made by the peptide amide to the phosphate backbone facilitates consolidation of the packing interactions in the region. (ii) Alternatively, strong contacts to DNA made simultaneously by the Arg 50 guanidino and amide groups might lead to a "conformational strain" and suboptimal packing of the side chain methylene groups to Ile 29 and Phe 33. If the second contact to DNA is removed by mutation, no structural changes are imposed in the region. It is not possible to distinguish between the two scenarios, yet both are fully compatible with the observed energetic signature of the R50A mutation. Loosening (suboptimal packing) of hydrophobic interactions in the HLH core is linked to disruption of enthalpically rich contacts, but a less tightly packed, and possibly more flexible region will be entropically favored. Also, although the difference is small in magnitude, $\Delta C_{p,\mathrm{A}}{}^{\mathrm{R50A}}$ is more negative than $\Delta C_{p,\mathrm{A}}{}^{\mathrm{WT}}$, as expected for a process linked to burial of hydrophobic surface. Not only for this site, a rigorous analysis of the energetic partitioning of the observed changes in affinity arising from structure perturbation is unfortunately hampered by the lack of high-resolution structural information on all relevant states comprising the thermodynamic system.

CONCLUSIONS

We have presented a complete description of the thermodynamic stability of the full length gene product of the Max transcription factor p21 isoform, and an in-depth characterization of the energetics of DNA binding. The combined results provide insights into the mechanistic basis of sequence-specific E-box recognition by proteins of the Myc/Max/Mad family. Direct heat capacity and enthalpy measurements demonstrate that the N- and C-terminal segments flanking the core b-HLH-LZ domain and accounting for half of the protein are disordered, do not participate in packing interactions, and are not involved in DNA binding. Their role in promoting the biological effects of Max is still poorly understood. However, the dissociation constant of Max p21 is in the low micromolar range (∼4 × 10^{-6} M) at 37 °C. At this temperature Max p21 binds the E-box target with affinity in the low nanomolar range. Considering that the nuclear concentration of Max p21 has been estimated to be about micromolar (*9*), the Max homodimeric state is populated and likely binds to E-box targets to affect transcription.

The DNA binding reaction is driven by net favorable enthalpy changes and is opposed by the net entropic changes. The refolding of the unstructured basic region of the free protein to the binding-competent α-helical state significantly contributes to the overall thermodynamic parameters. Stabilization of the LZ subdomain increases the affinity of Max p21 to DNA. The results suggest that thermodynamic coupling between HLH and LZ leads to subtle structural changes, as the consequence of which two Max residues form additional electrostatic contacts with the DNA phosphates, the likely candidates being Lys 30 from helix H1 and Lys 47 from the loop.

We have probed the energetic importance of evolutionary conserved residues that form well-defined contacts to the E-box target DNA bases or backbone phosphates according to X-ray data. Five side chains contribute modestly to binding affinity (15−160× decrease in K_D upon alanine mutation). One notable exception is Arg 26, which confers substantial stabilization. Although in close contact to the phosphate backbone, the Asn 19 side chain appears energetically unimportant. Altogether, the presented data point to the central role of the helical turn encompassing the C-terminus of the basic region and the start of helix H1. The preformed bivalent helical scaffold possibly anchors the protein to DNA at little expense of conformational entropy by positioning the energetically most important Arg 25 and Arg 26 to form three backbone contacts and one base-specific, deeply buried hydrogen bond. Association may be fast and dissociation may be slow since the clustered positive charges of Arg 25, Arg 26, and Arg 50 provide a steering force toward DNA. Furthermore, once stabilized, this conformation might serve as the nucleation site for propagation of α-helix toward the N-terminus of the basic region. Analysis of the thermodynamic signature of alanine mutants suggests that there are pronounced, context-dependent differences in the energetic content of formally identical protein−DNA contacts (for instance arginine−backbone phosphate bonds). Since the degree of sequence conservation in the basic and loop regions of Myc/Mad/Max family members is not absolute, future results from structural and biophysical experiments, and molecular mechanics approaches are required to elucidate the detailed structure−energetics relationships in DNA recognition by b-HLH-LZ proteins.

ACKNOWLEDGMENT

The authors thank Stoyan Milev for initial ITC experiments, Serge Chesnov for mass spectrometry analysis, and Christine Berger-Sprecher for stimulating discussions.

SUPPORTING INFORMATION AVAILABLE

Figures describing AUC analysis of the oligomerization state of Max p21, the influence of the monomer−dimer equilibrium on the shape of the binding isotherms, and analysis of the heat capacities of Max p21, E-box DNA, and the protein−DNA complex. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Amati, B., and Land, H. (1994) Myc−Max−Mad: a transcription factor network controlling cell cycle progression, differentiation and death, *Curr. Opin. Genet. Dev. 4*, 102−108.
2. Grandori, C., Cowley, S. M., James, L. P., and Eisenman, R. N. (2000) The Myc/Max/Mad network and the transcriptional control of cell behavior, *Annu. Rev. Cell Dev. Biol. 16*, 653−699.
3. Henriksson, M., and Luscher, B. (1996) Proteins of the Myc network: Essential regulators of cell growth and differentiation, *Adv. Cancer Res. 68*, 109−182.
4. Zhou, Z. Q., and Hurlin, P. J. (2001) The interplay between Mad and Myc in proliferation and differentiation, *Trends Cell Biol. 11*, S10−S14.
5. Toyo-oka, K., Bowen, T. J., Hirotsune, S., Li, Z., Jain, S., Ota, S., Lozach, L. E., Bassett, I. G., Lozach, J., Rosenfeld, M. G., Glass, C. K., Eisenman, R. N., Ren, B., Hurlin, P. J., and Wynshaw-Boris, A. (2006) Mnt-Deficient Mammary Glands Exhibit Impaired Involution and Tumors with Characteristics of Myc Overexpression, *Cancer Res. 66*, 5565−5573.
6. Banerjee, A., Hu, J., and Goss, D. J. (2006) Thermodynamics of Protein-Protein Interactions of cMyc, Max, and Mad: Effect of Polyions on Protein Dimerization, *Biochemistry 45*, 2333−2338.
7. Brownlie, P., Ceska, T. A., Lamers, M., Romier, C., Stier, G., Teo, H., and Suck, D. (1997) The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control, *Structure 5*, 509−520.
8. Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) Recognition by Max of Its Cognate DNA through a Dimeric B/Hlh/Z Domain, *Nature 363*, 38−45.
9. Nair, S. K., and Burley, S. K. (2003) X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors, *Cell 112*, 193−205.
10. Anthony-Cahill, S. J., Benfield, P. A., Fairman, R., Wasserman, Z. R., Brenner, S. L., Stafford, W. F., Altenbach, C., Hubbell, W. L., and Degrado, W. F. (1992) Molecular Characterization of Helix-Loop-Helix Peptides, *Science 255*, 979−983.
11. Fisher, D. E., Parent, L. A., and Sharp, P. A. (1993) High-Affinity DNA-Binding Myc Analogs - Recognition by an Alpha-Helix, *Cell 72*, 467−476.
12. Sauve, S., Tremblay, L., and Lavigne, P. (2004) The NMR solution structure of the max b/HLH/LZ free of DNA: Insights into the reversible DNA binding mechanism of dimeric transcription factors, *J. Mol. Biol. 13*, 161−162.
13. Hu, J. Z., Banerjee, A., and Goss, D. J. (2005) Assembly of b/HLH/z proteins c-Myc, Max, and Mad1 with cognate DNA: Importance of protein-protein and protein-DNA interactions, *Biochemistry 44*, 11855−11863.
14. Jung, K. C., Rhee, H. S., Park, C. H., and Yang, C.-H. (2005) Determination of the dissociation constants for recombinant c-Myc, Max, and DNA complexes: The inhibitory effect of linoleic acid on the DNA-binding step, *Biochem. Biophys. Res. Commun. 334*, 269−275.
15. Park, S., Chung, S., Kim, K.-M., Jung, K.-C., Park, C., Hahm, E.-R., and Yang, C.-H. (2004) Determination of binding constant of transcription factor myc-max/max-max and E-box DNA: the effect of inhibitors on the binding, *Biochim. Biophys. Acta 1670*, 217−228.
16. Naud, J., Montagne, M., McDuff, F., Chabot, B., and Lavigne, P. (2005) Structural and thermodynamical analysis of the complete p21 gene product of Max, *Biochemistry 44*, 12746−12758.
17. Naud, J.-F., Gagnon, F., Wellinger, R., Chabot, B., and Lavigne, P. (2003) Improving the Thermodynamic Stability of the Leucine Zipper of Max Increases the Stability of its b-HLH-LZ:E-box complex, *J. Mol. Biol. 326*, 1577−1595.
18. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves, *Methods Enzymol. 131*, 266−280.
19. Plotnikov, V. V., Brandts, J. M., Lin, L. N., and Brandts, J. F. (1997) A new ultrasensitive scanning calorimeter, *Anal. Biochem. 250*, 237−244.
20. Freire, E. (1995) Thermal denaturation methods in the study of protein folding, *Methods Enzymol. 259*, 144−169.
21. Privalov, P. L., and Potekhin, S. A. (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins, *Method Enzymol 131*, 4−51.
22. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid Measurement of Binding Constants and Heats of Binding Using a New Titration Calorimeter, *Anal. Biochem. 179*, 131−137.
23. Ferre-D'Amare, A. R., and Burley, S. K. (1994) Use of Dynamic Light-Scattering to Assess Crystallizability of Macromolecules and Macromolecular Assemblies, *Structure 2*, 567−567.

24. Manning, G. S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides, *Q. Rev. Biophys. 11*, 179−246.
25. Record, M. T., Jr., Ha, J. H., and Fisher, M. A. (1991) Analysis of equilibrium and kinetic measurements to determine thermodynamic origins of stability and specificity and mechanism of formation of site-specific complexes between proteins and helical DNA, *Methods Enzymol. 208*, 291−343.
26. Plotnikov, V., Rochalski, A., Brandts, M., Brandts, J. F., Williston, S., Trasca, V., and Lin, L.-N. (2002) An autosampling differential scanning calorimeter instrument for studying molecular interactions, *ASSAY Drug Dev. Technol. 1*, 83−90.
27. Richardson, J. M., Lopez, M. M., and Makhatadze, G. I. (2005) Enthalpy of helix-coil transition: Missing link in rationalizing the thermodynamics of helix-forming propensities of the amino acid residues, *Proc. Natl. Acad. Sci. U.S.A. 102*, 1413−1418.
28. Rose, G. D., Fleming, P. J., Banavar, J. R., and Maritan, A. (2006) A backbone-based theory of protein folding, *Proc. Natl. Acad. Sci. U.S.A. 103*, 16623−16633.
29. Dragan, A. I., Frank, L., Liu, Y. Y., Makeyeva, E. N., Crane-Robinson, C., and Privalov, P. L. (2004) Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites, *J. Mol. Biol. 343*, 865−878.
30. Doig, A. J., and Sternberg, M. J. E. (1995) Side-Chain Conformational Entropy in Protein-Folding, *Protein Sci. 4*, 2247−2251.
31. D'Aquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding, *Proteins 25*, 143−156.
32. Luque, I., and Freire, E. (1998) Structure-based prediction of binding affinities and molecular design of peptide ligands, *Method Enzymol. 295*, 100−127.
33. Baker, B. M., and Murphy, K. P. (1997) Dissecting the energetics of a protein-protein interaction: the binding of ovomucoid third domain to elastase, *J. Mol. Biol. 268*, 557−569.
34. Milev, S., Gorfe, A. A., Karshikoff, A., Clubb, R. T., Bosshard, H. R., and Jelesarov, I. (2003) Energetics of sequence-specific Protein-DNA association: Binding of integrase Tn916 to its target DNA, *Biochemistry 42*, 3481−3491.
35. Privalov, P. L., Jelesarov, I., Read, C. M., Dragan, A. I., and Crane-Robinson, C. (1999) The energetics of HMG box interactions with DNA: Thermodynamics of the DNA binding of the HMG box from mouse Sox-5, *J. Mol. Biol. 294*, 997−1013.
36. Olmsted, M. C., Bond, J. P., Anderson, C. F., and Record, M. T., Jr. (1995) Grand canonical Monte Carlo molecular and thermodynamic predictions of ion effects on binding of an oligocation (L8+) to the center of DNA oligomers, *Biophys. J. 68*, 634−647.
37. Bird, G. H., Lajmi, A. R., and Shin, J. A. (2002) Sequence-specific recognition of DNA by hydrophobic, alanine-rich mutants of the basic region/leucine zipper motif investigated by fluorescence anisotropy, *Biopolymers 65*, 10−20.
38. Park, S., Boder, E. T., and Saven, J. G. (2005) Modulating the DNA affinity of Elk-1 with computationally selected mutations, *J. Mol. Biol. 348*, 75−83.
39. Fisher, F., and Goding, C. R. (1992) Single Amino-Acid Substitutions Alter Helix Loop Helix Protein Specificity for Bases Flanking the Core Canntg Motif, *EMBO J. 11*, 4103−4109.
40. Dang, C. V., Dolde, C., Gillison, M. L., and Kato, G. J. (1992) Discrimination between Related DNA Sites by a Single Amino-Acid Residue of Myc-Related Basic Helix Loop Helix Proteins, *Proc. Natl. Acad. Sci. U.S.A. 89*, 599−602.
41. Sauve, S., Naud, J.-F., and Lavigne, P. (2007) The Mechanism of Discrimination between Cognate and Non-Specific DNA by Dimeric b/HLH/LZ Transcription Factors, *J. Mol. Biol. 365*, 1163−1175.
42. Griffith, K. L., and Wolf, J. R. E. (2002) A Comprehensive Alanine Scanning Mutagenesis of the Escherichia coli Transcriptional Activator SoxS: Identifying Amino Acids Important for DNA Binding and Transcription Activation, *J. Mol. Biol. 322*, 237−257.
43. Pace, C. N., and Scholtz, J. M. (1998) A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophys. J. 75*, 422−427.
44. Lacroix, E., and Viguera, A. R., & Serrano, L. (1998) Elucidating the folding problem of [alpha]-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol. 284*, 173−191.

BI701081Q

Supporting Information

Thermodynamics of b-HLH-LZ protein binding to DNA: The energetic importance of protein-DNA contacts in site-specific E-box recognition by the complete gene product of the Max p21 transcription factor

Laura Meier-Andrejszki, Saša Bjelić, Jean-François Naud., Pierre Lavigne and Ilian Jelesarov

Analytical ultracentrifugation analysis

Sedimentation equilibrium analysis was performed on a Beckman XL-A analytical ultracentrifuge. Protein and DNA solutions were dialyzed overnight against the standard buffer. The concentration of Max p21 was 10 μM and 60 μM (dimer equivalents), corresponding to the typical concentrations used in ITC and DSC experiments, respectively. The concentration of the complex was 1.5 μM. Since the binding constant at 25 °C is 3.8×10^8 M^{-1} (and is higher at 20 °C), more than 96 % protein/DNA complex was present in the cell. Equilibration was allowed for 40 hours at 20 °C. The data were sampled at rotor speeds of 18000, 20000 and 25000 rpm. Fifty scans (0.001 cm step size) were collected. After correction for the absorption of the reference cell (filled with plain buffer), the last 5 scans were averaged, and a single-species model was fitted to the data using the program WINNONLIN (http://spin6.mcb.uconn.edu/). Protein partial specific volume was calculated from the amino acid sequence. DNA partial specific volume was calculated with the program PSVOL (1) The partial specific volume of the complex was obtained from the weight fractions of protein and DNA in complex. In all experiments, the molecular mass best describing the data were all within 10 % of the expected molecular mass. More complicated models, i.e. formation of Max p21 tetramers, free or DNA-bound, did not improve the fitting statistics.

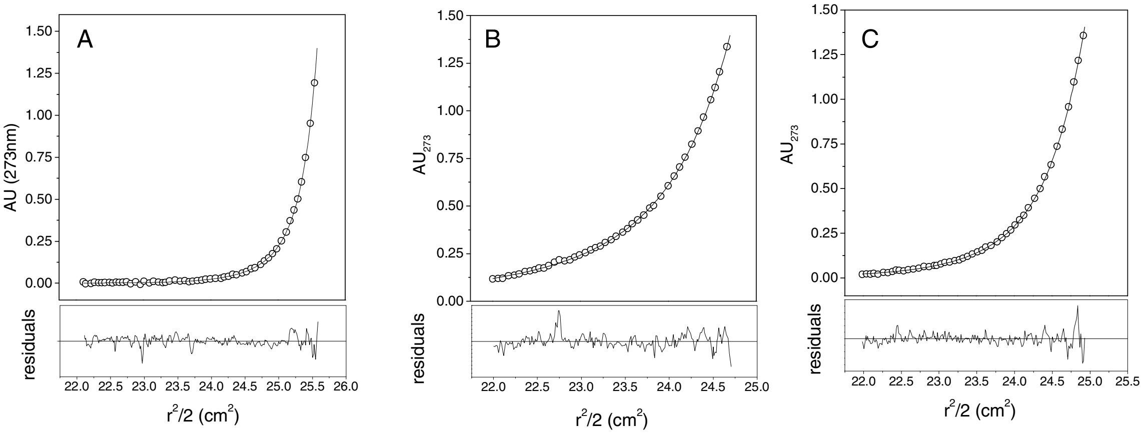


Figure S1. Analysis of the oligomerization state of Max p21 and the Max p21-DNA complex by sedimentation equilibrium analysis. Upper panels, UV absorbance gradients as function of the radial position (circles) and the fits according to a single species model (lines). Lower panels, Residuals showing the difference between the experimental data and the theoretical model. A, Max p21 at 10 μM concentration (dimer equivalents) and 18000 rpm. The fitted mass is 35.6 kDa. B, Max p21 at 60 μM concentration (dimer equivalents) and 25000 rpm. The fitted mass is 37.9 kDa. The calculated mass of the Max p21 protein is 17.07 kDa, i.e. the mass of the Max dimer is 34.1 kDa. C, protein-DNA complex at 1.5 μM concentration at 18000 rpm. The fitted mass is 48.8 kDa; the calculated mass is 47.0 kDa.

Influence of the Max p21 monomer-dimer equilibrium on the shape of calorimetric isoterms

In the following we present a simple model describing how the shape of the ITC isotherm changes when the monomer-dimer equilibrium of the Max p21 protein placed in the calorimetric cell steadily shifts due to formation of the dimer-DNA complex. For simplicity we assume that (i) all relevant equilibria are much faster than the response time of the instrument, (ii) there is no kinetic competition between the DNA binding reaction and protein folding/unfolding, and (iii) Max p21 monomers are unable to bind DNA, or else monomer binding is orders of magnitude weaker than dimer binding. At any given total concentrations of protein dimer (P_T) and DNA duplex (D_T) the equilibrium concentration of the protein-DNA complex (PD) can be calculated from the only physically meaningful root of the following quadratic equation:

$$K_A PD^2 + \left(P_T + D_T + \frac{1}{K_A}\right)PD + K_A P_T D_T = 0 \qquad \text{(S1)}$$

where K_A is the association constant. The free protein concentration (P_F; in dimer equivalents) in the cell (after correction for the dilution upon injectant addition) is

$$P_F = PT - PD \qquad \text{(S2)}$$

After each injection *i* the fraction of unfolded dimer is:

$$f_{U,i} = \frac{-K_D + \sqrt{K_D^2 + 4K_D P_F}}{2P_F} \qquad \text{(S3)}$$

K_D is the dissociation constant characterizing Max p21 monomer-dimer equilibrium. The concentration of unfolded free protein dimer after completion of the *i*-th injection is

$$P_{F,i}^{unf} = P_{F,i} f_{U,i} \qquad \text{(S4)}$$

The change in monomer-dimer equilibrium between consecutive injections is described by the change of P_F^{unf} between injections *i–1* and *i*:

$$\Delta P_{F,i}^{unf} = P_{F,i}^{unf} - P_{F,i-1}^{unf} = P_F \Delta f_U = P_F\left(f_{U,i} - f_{U,i-1}\right) \qquad \text{(S5)}$$

The heat associated with the shift of the dimer-monomer equilibrium is therefore:

$$\Delta q = q_i - q_{i-1} = \Delta P_F^{unf} \Delta H V_{cell} \qquad \text{(S6)}$$

where ΔH is the molar unfolding enthalpy of the dimer at the temperature of the ITC experiment and V_{cell} is the volume of the calorimetric cell. Figure S2 illustrates how the heat associated with the monomer-dimer equilibrium shift contributes to the total measured reaction heat. The figure presents the maximal possible distortion at the highest temperature (25 °C) and with concentrations in the lowest concentration range used in the ITC experiments. The effect is smaller at lower temperatures since the protein is more stable and, therefore both Δf_U and ΔH gets smaller as the temperatures decreases. It should be noted that the apparently negligible influence of the monomer-dimer equilibrium is a lucky consequence of the considered system. In cases where the intrinsic binding enthalpy is low, or the enthalpy of protein unfolding is large, or the protein is less stable, or any combination of these conditions takes place, the distortion of the binding isotherm could be pronounced.

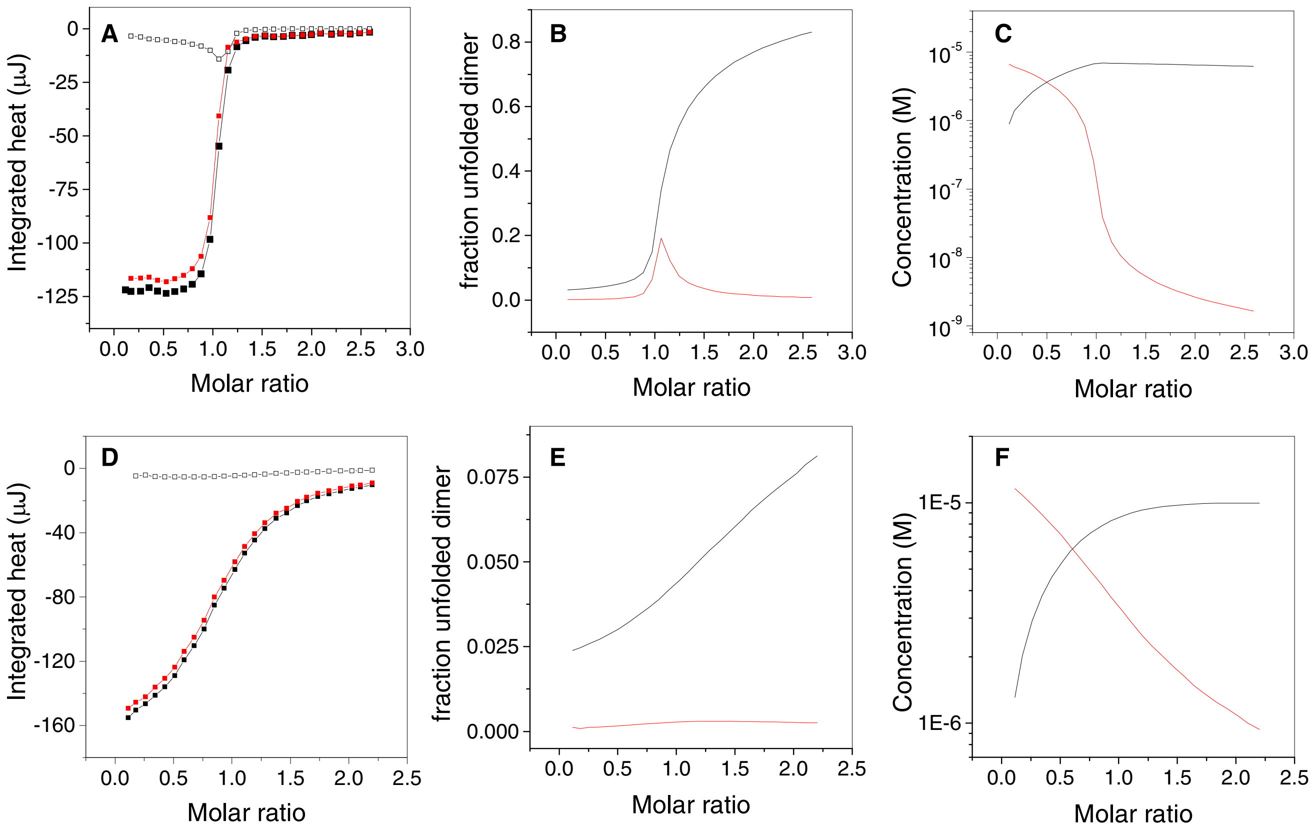


Figure S2. Distortion of the heat signal measured in when DNA is titrated into Max p21. The upper panels A- C represent high-affinity binding regime ($K_A = 3.8\times10^8$ M^{-1}, 7.9 µM Max dimer placed in the cell, 90 µM DNA in the injection syringe, c-value 3000). Lower panels D-F represent low-affinity binding regime ($K_A = 7.5\times10^5$ M^{-1}, 13 µM Max dimer placed in the cell, 150 µM DNA in the injection syringe, c-value 10). In the calculations the experimentally determined dissociation constant K_D(25 °C) = 2.7×10^{-8} M and unfolding enthalpy ΔH(25 °C) = 295 kJ mol^{-1} of Max p21 were used. In panels A and D, the black filled symbols are the experimentally measured integral heats at each injection. The open symbols are the heats associated with the shift of monomer-dimer equilibrium. The corrected heats (filled symbols minus open symbols) are shown in red. In panels B and D, the black line is the fraction free unfolded dimer at each injection, $f_{U,i}$. The red line is the difference $\Delta f_U = f_{U,i} - f_{U,i-1}$. In panels C and F, the increase of protein-DNA complex (black line) and the simultaneous decrease of free monomeric Max (red line) are shown.

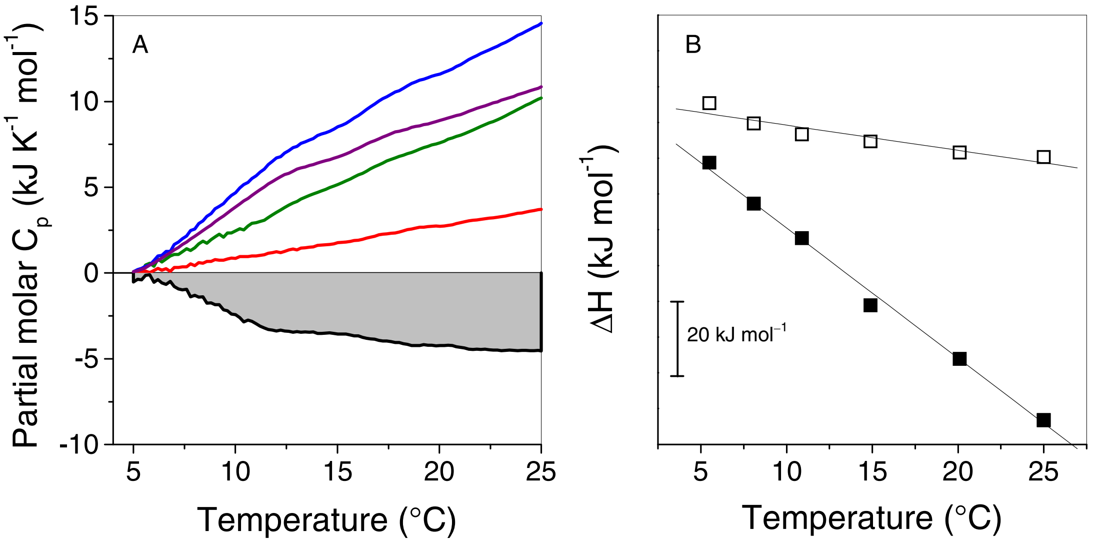


Figure S3. Estimation of the contribution of partial refolding and temperature-induced enthalpy fluctuations to the apparent measured $\Delta C_{p,A}$. Panel A, The temperature dependence of the partial molar heat capacities of free Max p21 (C_p^{MAX}; purple), free E-box duplex (C_p^{DNA}; red) and the 1:1 protein-DNA complex (C_p^{COMP}; green). Due to the limited accessible range of concentrations, the absolute heat capacities are not precisely known. For clarity, the traces are shifted to zero on the y-axis at 5 °C. The blue line represents the hypothetical heat capacity of the system in the dissociated state, calculated as the algebraic sum $C_p^{SUM} = C_p^{MAX} + C_p^{DNA}$ (blue = purple + red). At any temperature, the heat capacity change of association, taking the dissociated state as the reference state, is $\Delta C_{p,A} = C_p^{COMP} - C_p^{SUM} = C_p^{DIFF}$. The function C_p^{DIFF} is shown in panel A with the black line. The integral $\lambda = \int_{T_R}^{T} C_p^{DIFF} dT$ is the enthalpic contribution to ΔH_A arising from the fact that the heat capacities of the associated and dissociated states do not change in parallel upon temperature increase. The absolute value of $\lambda(T)$ can be evaluated at any arbitrary temperature only if $\lambda(T_R)$ is known. The latter function could in fact be regarded as the enthalpy of association at some temperature where there are no contributions to ΔH_A other than the enthalpy of formation of intermolecular bonds between "rigid-bodies" and the associated enthalpy of dehydration of groups, and changes of vibration. In the present case $\lambda(T_R)$ is not known. However, the relative values of $\lambda(T) \equiv$ can be calculated with reference to an arbitrarily selected T_R where $\lambda(T_R) = 0$. The shaded area in panel A represents $\lambda(T)^{rel}$ between 5 °C and 25 °C. Panel B shows the experimentally measured ΔH_A (filled symbols) and the function $\Delta H_A^{corrected} = \Delta H_A - \lambda(T)^{rel}$ (open symbols), arbitrarily shifted on the y-axis. The temperature dependence of ΔH_A is –3.8 kJ K^{-1} mol^{-1} and that of $\Delta H_A^{corrected}$ is –0.9 kJ K^{-1} mol^{-1}. The latter value could be taken as a rough estimate for the binding heat capacity change in the absence of basic region refolding and heat capacity effects associated with the unstructured segments of Max p21. The presented procedure is described in more details in ref. 2 and 3.

1. Karshikoff, A. & Ladenstein, R. (1998). Proteins from thermophilic and mesophilic organisms essentially do not differ in packing. *Protein Eng.* **11**, 867-872.
2. Privalov, P. L., Jelesarov, I., Read, C. M., Dragan, A. I. & Crane-Robinson, C. (1999). The energetics of HMG box interactions with DNA: Thermodynamics of the DNA binding of the HMG box from mouse Sox-5. *J. Mol. Biol.* **294**, 997-1013.
3. Milev, S., Gorfe, A. A., Karshikoff, A., Clubb, R. T., Bosshard, H. R. & Jelesarov, I. (2003). Energetics of sequence-specific Protein-DNA association: Binding of integrase Tn916 to its target DNA. *Biochemistry* **42**, 3481-3491.

List of Figures

ABBREVIATIONS

ASA, solvent accessible surface area
CD, circular dichroism
DSC, differential scanning calorimetry
C_p, heat capacity change
EB1c, C-terminal domain of end-binding protein 1
f_U, molar fraction of unfolded protein
ΔG_U, unfolding free energy
GdmCl, guanidinium hydrochloride
gp41, glycoprotein 41
ΔH, enthalpy change
ΔH_{vH}, enthalpy change ftom van't Hoff analysis
ITC, isothermal titration calorimetry
K_a, equilibrium association constant
k_f, refolding rate constant
k_u, unfolding rate constant
LEM, linear extrapolation method
Lpp-56, the protein moiety of the outer membrane lipoprotein of E. coli
m_{rq}, the first derivative of the observed unfolding free energy with respect to the denaturant concentration
m_f and m_u, the first derivatives of the logarithm of the folding rate constant and unfolding rate constant, respectively, with respect to the denaturant concentration
MRE_{225}, mean residue ellipticity at 225 nm ΔS, entropy change.

ACKNOWLEDGMENTS

- I am especially thankful to PD Dr. Ilian Jelesarov allowing me to participate in exiting biophysical research. I must emphasize numerous open and inspiriting discussions as well as freedom and almost endless support in exploring ...
- To Prof. Amedeo Caflish for being my doctoral thesis supervisor and helpful discussions on analysis of molecular dynamic simulations.
- Prof. N. N. for efforts associated with PhD examen. ??
- To Prof. Andrey Karshikoff for excellent collaboration.
- To Dr. Daniel Nettels for help in Mathematica scripting.
- To Dr. Rainer Bckmann for introducing me in the world of molecular dynamic simulations and useful discussions.
- To Philipp Schuetz and Dr. Raffaele Curcio for help scripting for molecular dynamic analysis.
- To Dr. Stephan Klauser, Steve Rast and Thomas Hirt as well Matterhorn-cluster administrators for all kind of IT-related problems.
- To my former and present colleagues from PD Dr. Ilian Jelezarov's and Prof. Bosshards group thank you for your help and stimulating atmosphere.
- To all former and present members of Biochemistry Institute for direct and indirect help in everyday life.
- Last but not list I have to express my gratitude to my family for all kind of support and many hours they have to be alone.

CURRICULUM VITAE

Saša BJELIĆ

Personal details

Date of Birth	April 15, 1977
Nationality	Swiss, Wilchingen SH
Marital status	married, 2 children (4 and 1 years)

Education and Employment

2004 - present	Doctoral studies at the Department of Biochemistry, University of Zurich, PhD thesis: "Folding and Stability of Homooligomeric Coiled coils"
2004	M. Sc. in Biochemistry; M.Sc. Thesis: "Folding Thermodynamics and Kinetics of Surface Protein gp41 from HI-Virus"
1999 - 2004	Biochemistry, University of Zurich, Zurich, Switzerland
1995 - 1999	High School AKAD, Zurich, Switzerland

Teaching

2006	Bachelor work: Electrostatic Contribution on Stability of Microtubule Protein Eb1c

2006	Research-oriented practicum II
2005	Bachelor work: Folding and Stability of a gp41 Mutant
2005	Research-oriented practicum II
2004	Research-oriented practicum II
2004	Research-oriented practicum I

List of publications

- Marti, D. N, Bjelić, S, Lu, M, Bosshard, H. R, & Jelesarov, I. (2004) Fast folding of the HIV-1 and SIV gp41 six-helix bundles. *J. Mol. Biol.* **336**, 1–8.
- Bjelić, S, Karshikoff, A, & Jelesarov, I. (2006) Stability and folding/unfolding kinetics of the homotrimeric coiled coil Lpp-56. *Biochemistry* **45**, 8931–8939.
- Bjelić, S, Wieninger, S, Jelesarov, I, & Karshikoff, A. (2007) Electrostatic contribution to the thermodynamic and kinetic stability of the homotrimeric coiled coil Lpp-56: A computational study. *Proteins* **70**, 810–822.
- Milev, S, Bjelić, S, Georgiev, O, & Jelesarov, I. (2007) Energetics of peptide recognition by the second PDZ domain of human protein tyrosine phosphatase 1E. *Biochemistry* **46**, 1064–1078.
- Meier-Andrejszki, L, Bjelić, S, Naud, J.-F, Lavigne, P, & Jelesarov, I. (2007) Thermodynamics of b-HLH-LZ protein binding to DNA: the energetic importance of protein-DNA contacts in site-specific E-box recognition by the complete gene product of the Max p21 transcription factor. *Biochemistry* **46**, 12427–12440.